SSTC 2010

# *A Tutorial for Building CMMI Process Performance Models*

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA  15213

Robert Stoddard and Dave Zubrow
April 26, 2010

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **26 APR 2010** | | **00-00-2010 to 00-00-2010** |

4. TITLE AND SUBTITLE
**A Tutorial for Building CMMI Process Performance Moels**

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
**Carnegie Mellon University,Software Engineering Institute,Pittsburgh,PA,15213**

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**Presented at the 22nd Systems and Software Technology Conference (SSTC), 26-29 April 2010, Salt Lake City, UT. Sponsored in part by the USAF. U.S. Government or Federal Rights License**

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | |
| **unclassified** | **unclassified** | **unclassified** | **Same as Report (SAR)** | **200** | |

# Permissions (Crystal Ball and JMP Screen Shots)

Portions of the input and output contained in this module manual are printed with permission of Oracle (formerly Decisioneering). Crystal Ball 7.2.2 (Build 7.2.1333.0) is used to capture screenshots in this module

The Web page for Crystal Ball is available at http://www.crystalball.com

---

Screen shots and other statistical tool information has been used with permission from SAS Institute. Information about JMP® statistical discovery software can be found at www.jmp.com.

JMP® is interactive, comprehensive, visual software from SAS. It dynamically links statistics with graphics right on your Windows, Macintosh, or Linux desktop, empowering you to explore data interactively and bring understanding to your organization.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies. Copyright © 2007 SAS Institute Inc. All rights reserved. 449113.0607

# Acknowledgment

This tutorial reuses portions of a full day tutorial given at the 2009 SEPG North America, but with 5 hands-on, guided exercises added to enable participant skill development with process performance models.

Appropriate acknowledgment is made to the authors of the 2009 SEPG NA tutorial:

Kevin Schaaff, Robert Stoddard, Rusty Young and Dave Zubrow

# Topics

Introduction (***10 min***)

Overview of the Steps to Build PPMs (***80 mins***)

- Preparing to Build PPMs
- Developing PPMs
- Using PPMs

Exercise 1: Constructing a Product Business Case with Monte Carlo Simulation and Optimization (***40 mins***)

Exercise 2: Scheduling Projects with Monte Carlo Simulation and Optimization (***30 mins***)

Exercise 3: Predicting Product Requirements Change with Linear Regression (***30 mins***)

Exercise 4: Predicting Delivered Defects with Dummy Variable Regression (***30 mins***)

Exercise 5: Predicting Customer Satisfaction using Ordinal Logistic Regression

Questions (***30 mins***)

# Introduction

# What is a PPM?

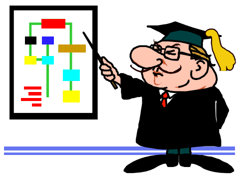
## OPP SP 1.5

- **PPMs** are used to estimate or predict the value of a process-performance measure from the values of other process, product, and service measurements
- **PPMs** typically use process and product measurements collected throughout the life of the project to estimate progress toward achieving objectives that cannot be measured until later in the project's life

## Glossary

- A description of the relationships among attributes of a process and its work products that is developed from historical process-performance data and calibrated using collected process and product measures from the project and that is used to predict results to be achieved by following a process

# Purpose and Usage of Process Performance Models at the Organizational Level

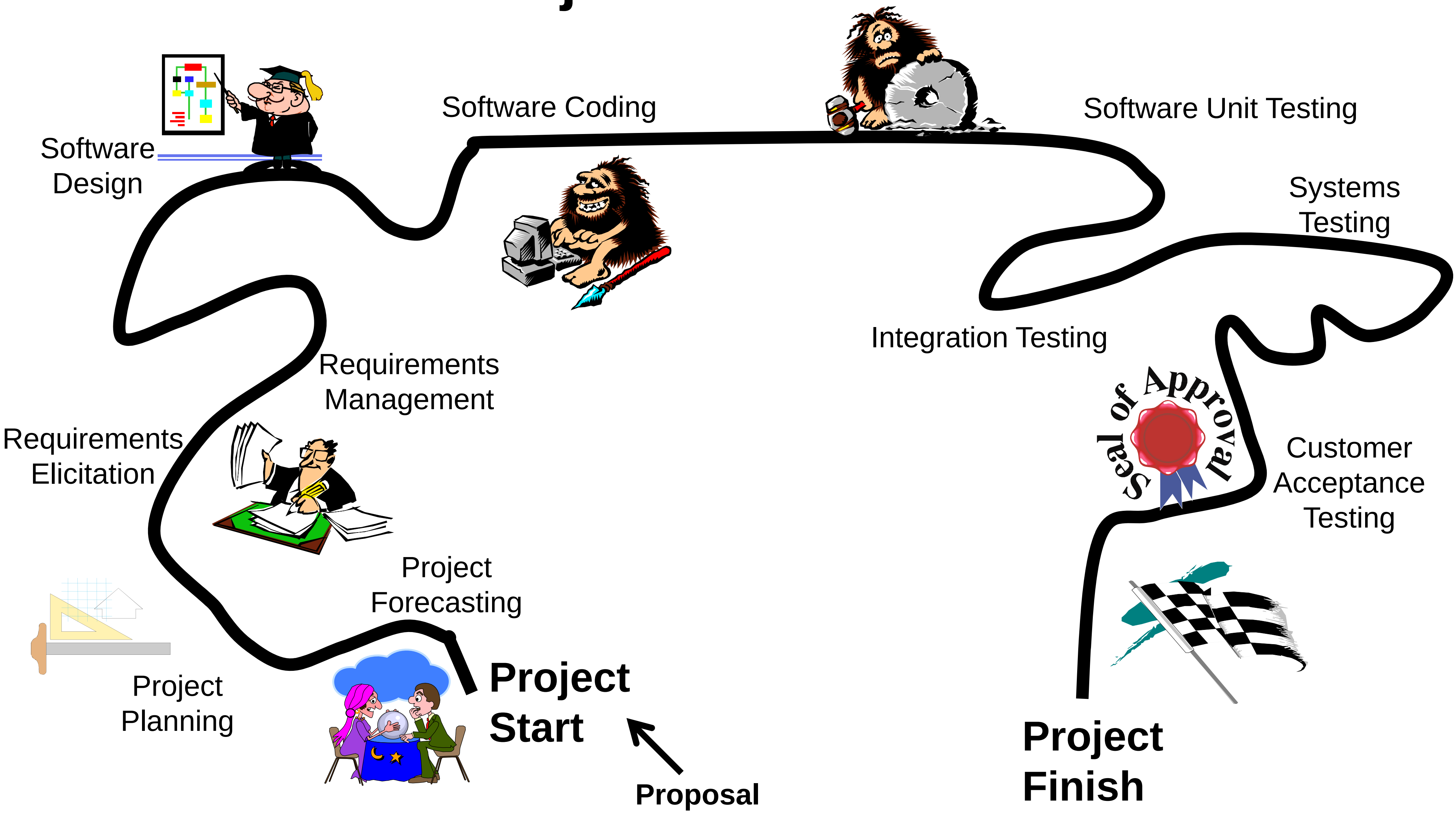


- Identifying Organizational Priorities for Quality and Process Performance
- Establishing and Revising Organizational Quality and Process Performance Objections

- Identifying Process Performance Measures
- Defining New Process Performance Baselines

- Analyzing Process and Technology Improvement Proposals
- Identifying Process and Technology Improvement Proposals
- Prioritizing Candidate Process and Technology Improvements for Deployment

# Purpose and Usage of Process Performance Models at the Project Level



# Healthy Ingredients of CMMI Process Performance Models

Statistical, probabilistic or simulation in nature

Predict interim and/or final project outcomes

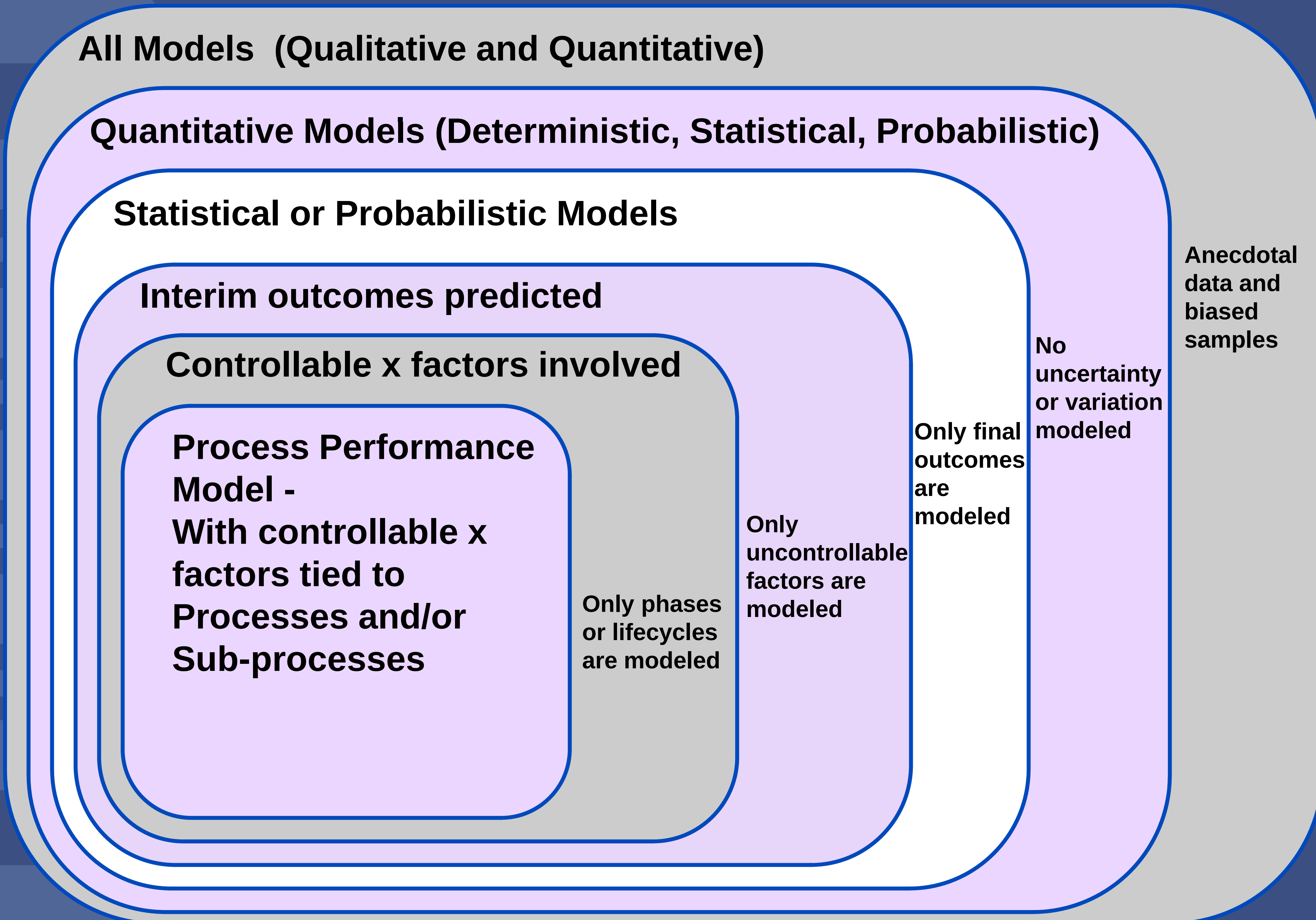
Use controllable factors tied to sub-processes to conduct the prediction

Model the variation of factors and understand the predicted range or variation of the outcomes

Enable “what-if” analysis for project planning, dynamic re-planning and problem resolution during project execution

Connect “upstream” activity with “downstream” activity

Enable projects to achieve mid-course corrections to ensure project success


All Models (Qualitative and Quantitative)
Quantitative Models (Deterministic, Statistical, Probabilistic)
Statistical or Probabilistic Models
Interim outcomes predicted
Controllable x factors involved
Process Performance Model - With controllable x factors tied to Processes and/or Sub-processes
Only phases or lifecycles are modeled
Only uncontrollable factors are modeled
Only final outcomes are modeled
No uncertainty or variation modeled
Anecdotal data and biased samples

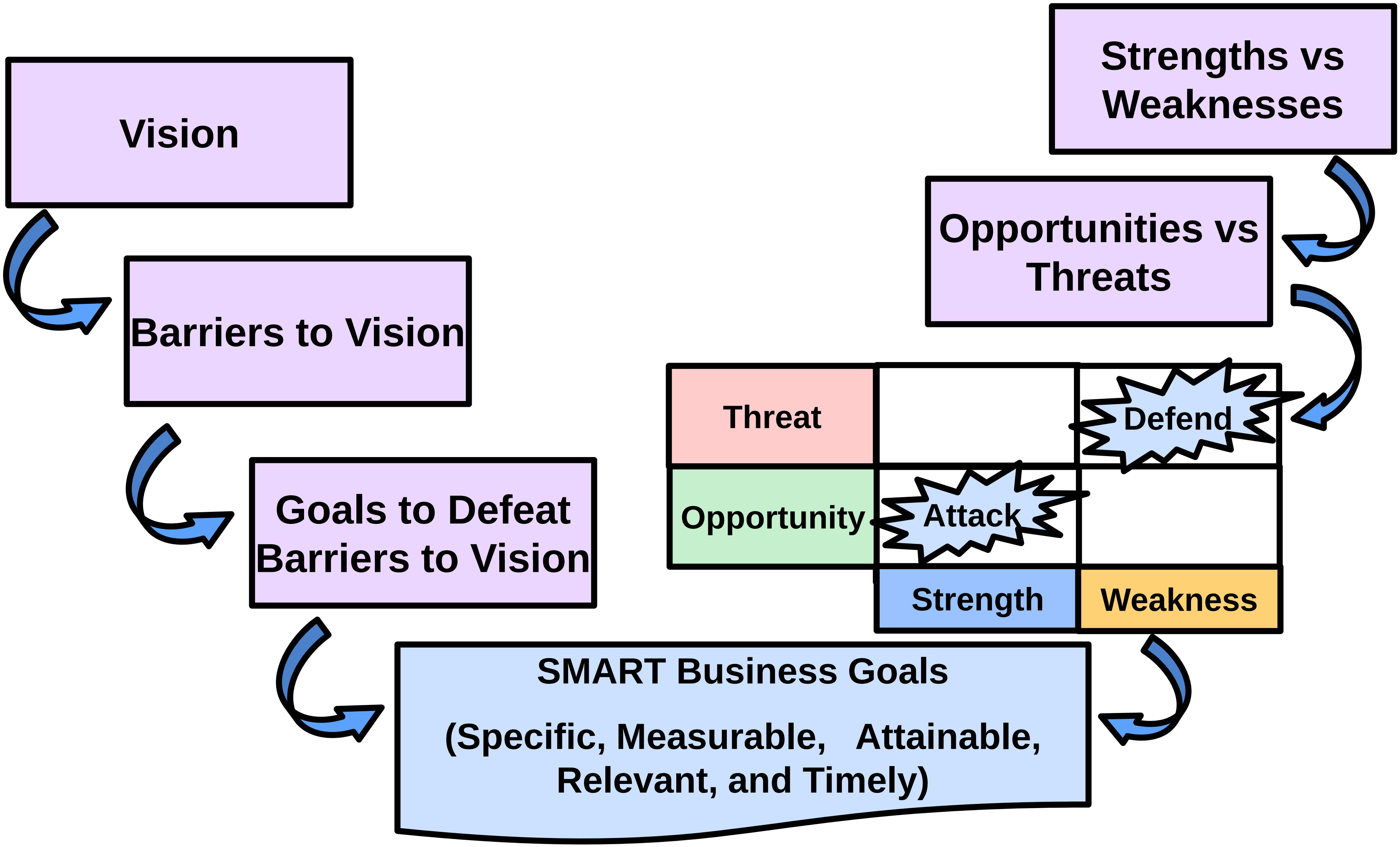
# Overview of the Steps to Build PPMs

## - Preparing to Develop PPMs

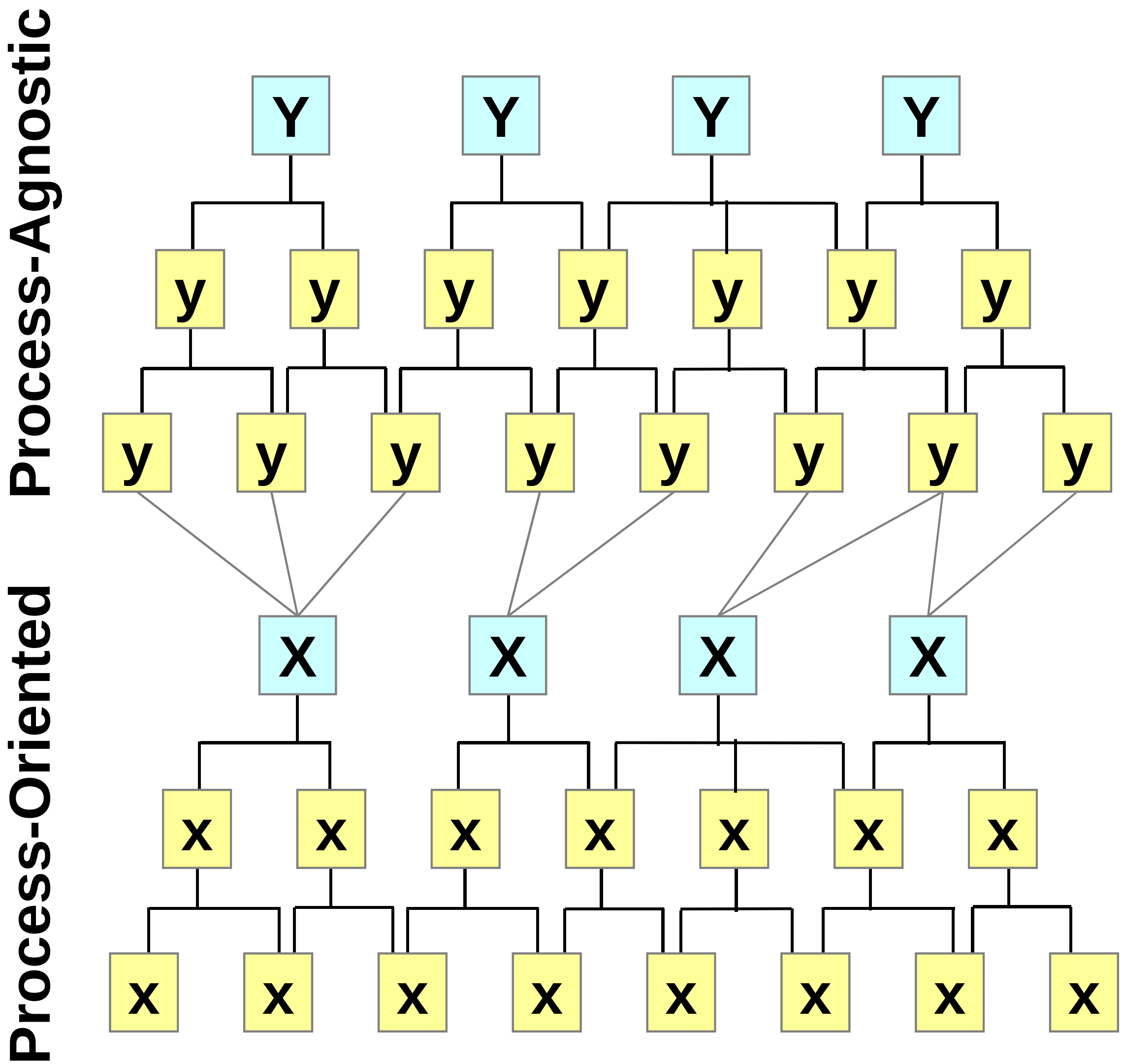
# Preparing to Develop PPMs

1. Initiating the development of process performance models from a context of the customer and Business goals
2. Using correct critical thinking and root cause analysis to identify the proper outcomes and drivers of the outcomes (including controllable and uncontrollable process factors)
3. Becoming sensitive to the types of issues and documentation needed during the development of the process performance models
4. Addressing issues related to data collection, measurement scale, data quality and integrity, outliers and measurement error
5. Identifying the data types involved with the outcomes and process drivers
6. Creating performance baselines of outcomes and process drivers
7. Forming a team to develop a process performance model

# Step 1 - Identify or Reconfirm Business Goals



# Step 1 - Business Goal Flowdown (Y-to-x)




**High Level Business Goals**
**(Balanced Scorecard)**

**Subordinate Business Goals**
**(e.g., $ Buckets, % Performance)**

**High Level Process**
**(e.g., Organizational Processes)**

**Subordinate Processes**
**(e.g., Down to a Vital x sub-process to be tackled by DMAIC team)**

# Questions

1. Are your senior leaders defining business goals rather than delegating goal definition to operational levels?

2. Do lower organizational levels redefine the higher level goals in operational terms or do they merely block copy and paste upper goals?

3. Are you organization's business goals SMART?

4. Has your organization ensured that process performance baselines and models are targeted at the most important issues and goals?

# Step 2 - Identify the Sub-Process/Process

- Start with the Organization's Business Objectives
- Decompose to Quality and Process Performance Objectives (QPPOs)
- For the QPPOs that can be Measured Quantitatively
  - Perform Analysis to Determine which Sub-Process/Process Drives the Relevant Objective
  - Determine if Sufficient Data is Available or can be Obtained to Establish a Process Performance Baseline(s) and/or Build a Process Performance Model(s)

# Step 2 - Identify the Sub-Process/Process Example

- Given Organizational Business Objectives:
  - Improve quality
  - Improve cycle time
  - Improve productivity
- Translate to measureable QPPOs
  - Post-delivery defect density of less than 0.5 Defects/KSLOC
  - Achieve 85% defect detection before System testing
  - Ensure requirements duration is within 15% of plan
  - Achieve a 5 % software productivity improvement

# Step 2 - Examples of Outcomes

Injected Defects Volume by type

Escaped defects by phase*

Task duration

Availability of resources*

Task delay

Schedule Variance

Task effort

Cost Variance

Earned Value Metrics (CPI, SPI)

Latent defect content of artifact*

Difficulty*

Productivity*

Rework

Req’ts Volatility*

Cost of Poor Quality

Customer Satisfaction

Time to Market

Progress*

“ilities” such as Reliability

Warranty Costs



Robert Stoddard and Dave Zubrow

# Step 2 - Identify Controllable factors (x's) to Predict Outcome(s) - 1

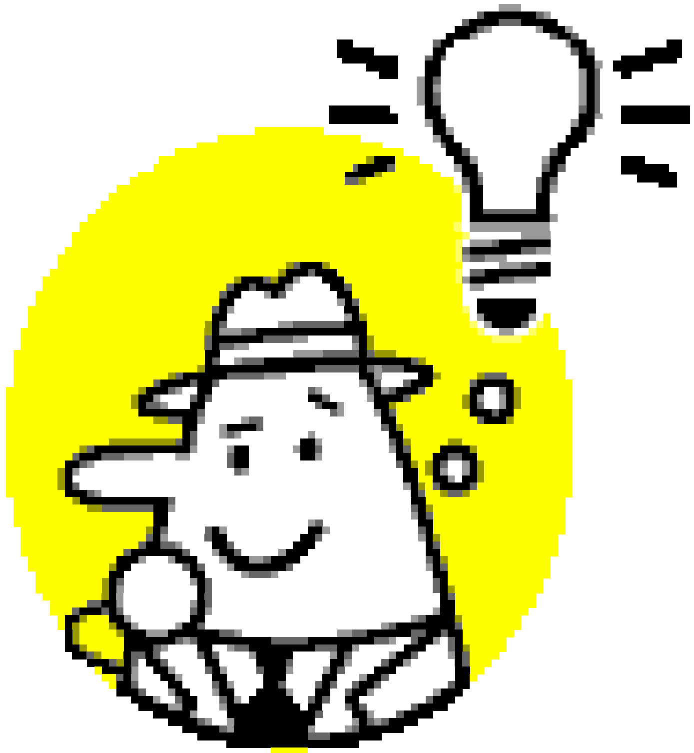
"Controllable" implies that a project has direct or indirect influence over the factor prior to or during the project execution

A common misconception is that factors are not controllable and thus disregarded from consideration for modeling. Requires out-of-the-box thinking to overcome this. Some organizations employ individuals known as "assumption busters"

# Step 2 - Identify Controllable factors (x's) to Predict Outcome(s) - 2

As we view process holistically, controllable factors may be related, but not limited, to any of the following:

- People attributes
- Environmental factors
- Technology factors
- Tools (physical or software)
- Process factors
- Customers
- Suppliers
- Other Stakeholders



# Step 2 - Examples of Controllable People x factors

Absolute performance of a task or topic

Training

Variability of performance of a task or topic

Skills

Interruptions

Degree of Mentoring and Coaching

Traits

Degree of Multi-tasking

Staff Availability

Experience Levels

Geographic dispersion of staff

Diversity of staff

Attitudes and Outlooks

Communication Mechanisms

Various Teaming Attributes

Knowledge Sharing Mechanisms

Degree of Cross Training

Multi-capable staff

Organizational Dynamics

Nature of Leadership

# Step 2 - Example of Controllable Environmental x Factors

Nature of work facilities

Access to breakout rooms

Degree of noise or distractions

External interferences including other organizations

Proximity to team members

Access or proximity to customers

Ergonomics

Temperature

Accomodations for specific needs

Access or proximity to suppliers

Available Training Rooms

Access or proximity to management and other stakeholders

Degree of Security Classification

Other Visual or Audio Distractions



Robert Stoddard and Dave Zubrow

# Step 2 - Example of Controllable Technology x Factors

Degree of modern development tools

Mature tools

Degree technology proven

Newness of Technology

Availability of equipment, test stations

Availability of Technology

Complexity of Technology

Documentation of Technology

Newness of Technology

Programming Language Used

Competition use of technology

Platform or Operating System Used

Technology Trends

Nature of Legacy or Reuse

Technology Roadmap

# Step 2 - Example of Controllable Process x Factors

**Resolution time of technical inquiries**

**Efficiency of a work task**

**Compliance of a work task**

**Quality of a work task**

**Timeliness of a work task**

**Measures of bureaucracy**

**Resource contention between tasks**

**Difficulty of a work task**

**Number of people involved with a work task**

**Degree of Job Aids, Templates, Instructions**

**Choices of subprocesses**

**Peer Review Measures**

**Test Coverage Measures**

**Modifications to how work Tasks are performed**

**Quality of artifacts (Input to or Output from a work task)**

**Timeliness of Artifacts**

**Task Interdependence**

**Complexity of Artifacts**

**Readability of Artifacts**

**Any of the criteria for good reqts statements**

**Any of the criteria for good designs**

**Code measures (Static and Dynamic)**

# Step 2 - Example of Controllable Customer, Supplier and Other Stakeholder x Factors

"Maturity" assessment

Volatility of Staff

Early Involvement

Conflicts among Stakeholders

Health of relationship

Degree of Documentation of Expectations

Degree of communication

Image and Perceptions

Speed of feedback loops

Complexity of relationship such as simultaneously a competitor and partner and supplier

Trust

Longevity of relationship

Degree of oversight

Style

Degree of partnership, collaboration

Bias on Quality vs Schedule

Geographic location

Culture

Domain Experience

Language

Degree of access and participation

Tradeoffs, Compromises, Optimization

# Step 2 - Identify Uncontrollable Factors

- Normally these are constraints placed by the customer or concrete terms of a contract or government regulation

- Can also be factors for which the project team truly has no direct nor indirect influence over

- Can be factors that are unchanging for a given project but can be changed for future projects

- Often includes external factors or factors related to other teams outside of the project

# Questions

1. What is a critical, high risk, uncertain subprocess within your organization?

2. What is a potential outcome performance measure related to that subprocess?

3. What are 2-3 controllable factors directly influencing this outcome measure?

4. Do you believe there are any uncontrollable factors dominating this outcome measure?

# Step 3 - Cost of Poor Data Quality to an Enterprise – Typical Issues and Impacts

## Typical Issues

- Inaccurate data [1-5% of data fields are erred]
- Inconsistencies across databases
- Unavailable data necessary for certain operations or decisions

## Typical Impacts

| Operational | Tactical | Strategic |
|---|---|---|
| • Lowered customer satisfaction<br>• Increased cost<br>• Lowered employee satisfaction | • Poorer decision making & decisions take longer<br>• More difficult to implement data warehouses<br>• More difficult to engineer<br>• Increased organizational mistrust | • More difficult to set strategy<br>• More difficult to execute strategy<br>• Contribute to issues of data ownership<br>• Compromise ability to align organization<br>• Divert management attention |

**Source: Redman, 1998**

# Step 3 - Impacts of Poor Data Quality

Inability to

- manage the quality and performance of software or application development
- Estimate and plan realistically

Ineffective

- process change instead of process improvement
- and inefficient testing causing issues with time to market, field quality and development costs

Products that are painful and costly to use within real-life usage profiles

**Bad Information leading to Bad Decisions**

# Step 3 - Where do Measurement Errors come From[1]

Data Entry Errors

- Manual data entry
- Lack of integrity checks

Differing Operational Definitions

- Project duration, defect severity or type, LOC definition, milestone completion

Not a priority for those generating or collecting data

- Complete the effort time sheet at the end of the month
- Inaccurate measurement at the source

Double Duty

- Effort data collection is for Accounting not Project Management
  - Overtime is not tracked
  - Effort is tracked only to highest level of WBS

# Step 3 - Where do Measurement Errors come From₂

## Dysfunctional Incentives

- Rewards for high productivity measured as LoC/Hr
- Dilbert-esque scenarios

## Failure to provide resources and training

- Assume data collectors all understand goals and purpose
- Arduous manual tasks instead of automation

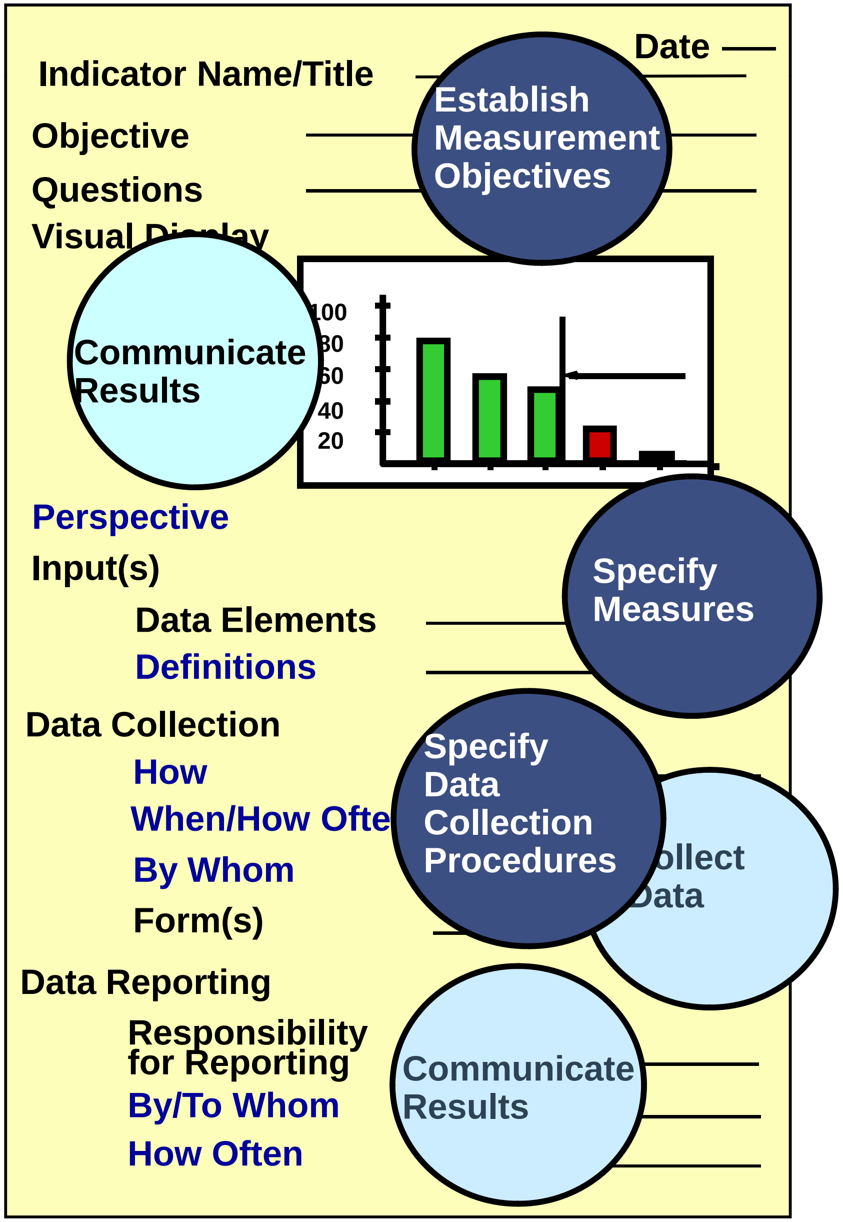
## Lack of priority or interest

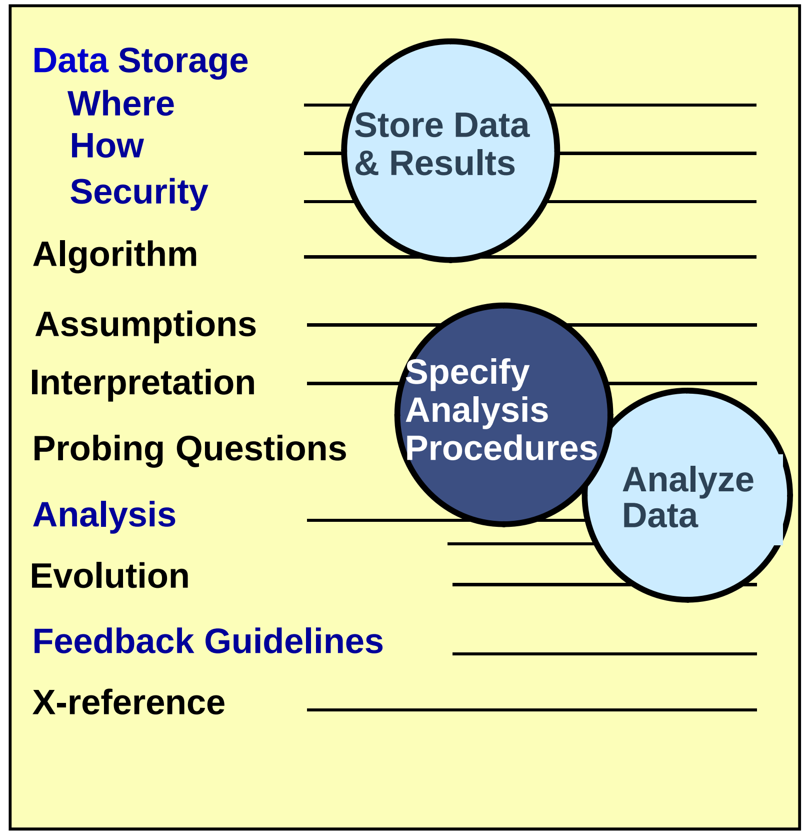
- No visible use or consequences associated with poor data collection or measurement
- No sustained management sponsorship

## Missing data is reported as a valid value

- Can’t distinguish 0 from missing when performing calculations

# Step 3 - Documenting Measurement Objectives, Indicators, and Measures



# Step 4 - Identifying Outliers

Interquartile range description – A quantitative method for identifying possible outliers in a data set

Procedure

- Determine 1st and 3rd quartiles of data set: Q1, Q3
- Calculate the difference: interquartile range or IQR which equals Q3 minus Q1
- Lower outlier boundary = Q1 – 1.5*IQR
- Upper outlier boundary = Q3 + 1.5*IQR

# Step 4 - Interquartile Range: Example

2

Interquartile Range
30 – 16 = 14

| | |
|---|---|
| | 333 |
| 1 | 50 |
| | 40 |
| Q3 → | 30 |
| | 27 |
| | 25 |
| | 22 |
| | 20 |
| | 18 |
| Q1 → | 16 |
| | 16 |
| | 13 |

4

Upper outlier boundary
30 + 1.5*14 = 51

3

Lower outlier boundary
16 – 1.5*14 = -5

**Procedure**

1. **Determine 1st and 3rd quartiles of data set: Q1, Q3**
2. **Calculate the difference: interquartile range or IQR**
3. **Lower outlier boundary = Q1 – 1.5*IQR**
4. **Upper outlier boundary = Q3 + 1.5*IQR**

# Step 4 - Tips About Outliers

Outliers can be a clue to process understanding

If outliers lead you to measurement system problems,

- repair the erroneous data if possible
- if it cannot be repaired, delete it

Charts that are particularly effective to flag possible outliers include:  box plots, distributions, scatter plots, and control charts

Rescale charts when an outlier reduces visibility into variation.

**Be wary of influence of outliers on linear relationships**

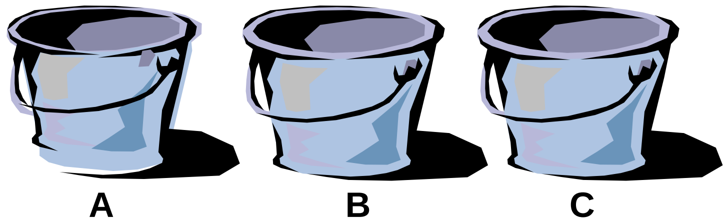
SIX SIGMA ADVANTAGE
The Third Wave™

# Step 5 - Types of Data

**Attribute**
(aka categorized or discrete data)

**Nominal**

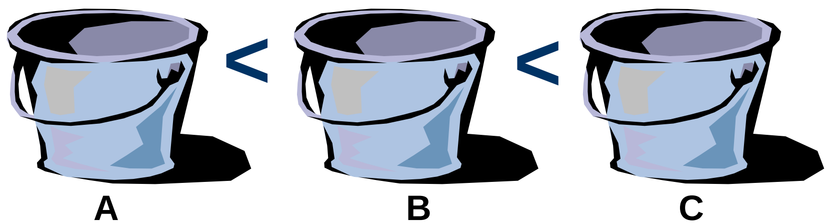
Categorical data where the order of the categories is arbitrary




**Examples**
Defect types
Labor types
Languages

**Ordinal**

Nominal data with an ordering; may have unequal intervals




**Examples**
Severity levels
Survey choices 1-5
Experience categories

*Increasing information content*

---

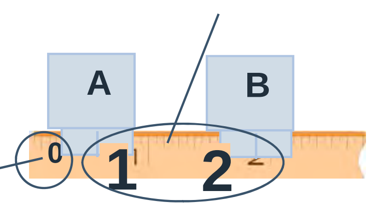
**Continuous**
(aka variables data)

**Interval**

Continuous data that has equal intervals; may have decimal values

**Ratio**

Interval data set that also has a true zero point




**Examples**
Defect densities
Labor rates
Productivity
Variance %'s
Code size SLOC

# Questions

1. What data type is your outcome performance measure?

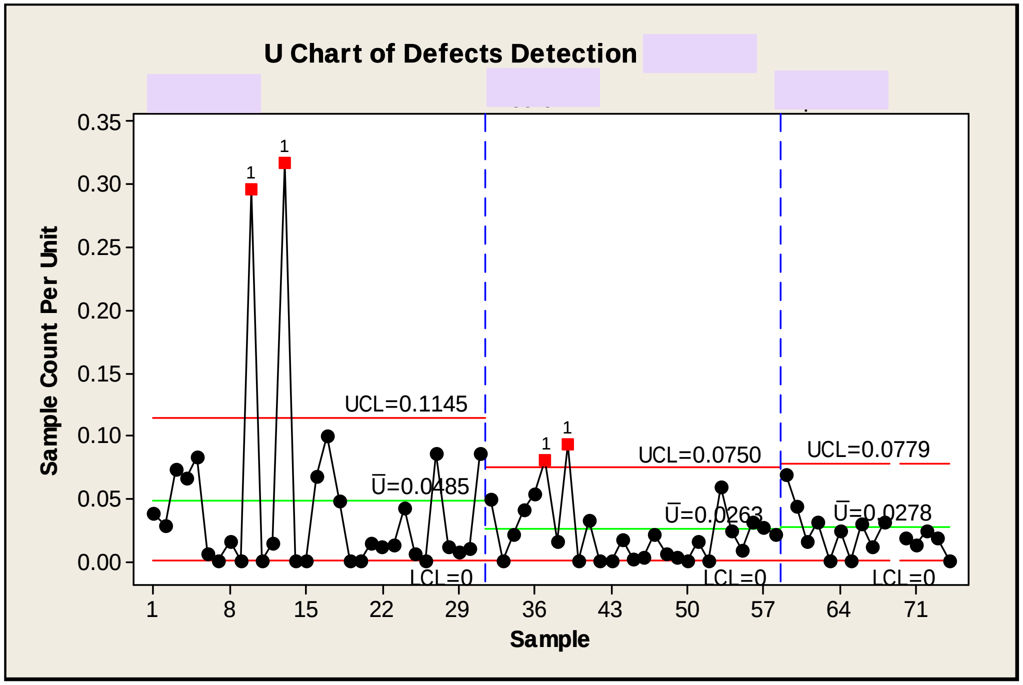
2. What data type is each of your controllable and uncontrollable x factors?

# Step 6 - Creating Process Performance Baselines

- Definition: A Process Performance Baselines (PPB) is a documented characterization of the actual results achieved by following a process
- Therefore a PPB needs to reflect actual project performance
- CMMI-DEV OPP PA informative material:
  - Establish a quantitative understanding of the performance of the organization's set of standard processes in support of objectives
  - Select the processes that summarize the actual performance of processes in projects in the organization
- Alternatively Practical Software and Systems Measurement (PSM) recommends an organization follow three basic steps:
  - Identify organization needs
  - Select appropriate measures
  - Integrate measurement into the process

# Step 6 - Creating Process Performance Baselines Example

- If we go back to our earlier example where we determined that the inspection sub-process should be statistically managed
- Collect data and Establish a PPB for the inspection sub-process



# Step 6 - Appropriate Analysis: Types of Hypothesis Tests

| Data Type → / # Samples (Data groups) ↓ | Interval or Ratio (Parametric Tests): Mean | Interval or Ratio (Parametric Tests): Variance | Ordinal (Non-Parametric Tests): Median | Ordinal (Non-Parametric Tests): Variance / Fit | Nominal: Similarity | Proportion: Similarity |
|---|---|---|---|---|---|---|
| 1 Sample | 1-sample t test | 1-sample Chi-Square test | 1 sample Wilcoxon Signed Ranks test | Kolmogorov-Smirnov Goodness of Fit test | *>2 cells*: Chi-Square / *=2 cells*: Binomial Sign Test | 1 Proportions test |
| 2 Samples | *Independent*: 2-sample t test / *Paired*: Paired t test | *Normal*: F test / *Not Normal*: Levene test | *Independent*: Mann Whitney U test / *Paired*: Wilcoxon matched | *= Medians*: Siegel-Tukey test / *≠ Medians*: Moses test | Fisher Exact test (1-way ANOVA); Chi-Square test | 2 Proportions test |
| 3+ Samples | ANOVA (1 & 2 way ANOVA; Balanced ANOVA; GLM) MANOVA (General & Balanced) | *Normal*: Bartlett test / *Not Normal*: Levene test | *Independent*: Kruskal-Wallis 1-way ANOVA / *Paired*: Friedman 2-way ANOVA | Van der Waerden Normal scores test | Chi-Square test | ANOM (Analysis of Means) |

# Step 6 - Creating Process Performance Baselines Misconceptions

- We only need one baseline
- Once we establish the initial set of baselines we are done
- One data point constitutes a baseline
- We can’t use the baseline until it is stable
- If the initial baseline is unstable we just remove the data points outside of the control limits and recompute the control limits until we get a plot that appears stable

# Step 7 - Skills Needed to Develop PPMs

- Business Acumen
- Product Expertise
- Process Expertise
- Understanding of Measurement and AnalysisTechniques
- Understanding of Advanced Statistical Techniques
- Understanding of Quantitative Management

# Step 7 - Forming the PPM Development Team

## Statistical Skills

- PPM builder needs a good understanding of statistics or Six Sigma Black Belt skill level or better
- PPM builder needs to be an expert user of the selected statistical tools
- User of PPMs needs to be an educated consumer

## Process knowledge

- Build team needs to understand the process
- Build team needs to understand the context in which the PPMs will be used

# Overview of the Steps to Build PPMs

## - Creating PPMs

# Creating PPMs

1. Identifying and using the correct analytical techniques for analyzing baselines, and creating process performance models
2. Creating both confidence and prediction intervals with the models
3. Validating and maintaining the process performance models including calibration and re-confirming with ongoing process and project data
4. Confirming process performance models meet the established ingredients communicated by the SEI, either individually or as a whole

# Step 1 - Select the Proper Analytical Model

## Types of Modeling Techniques

- Statistical Modeling and Regression Equations
- Monte Carlo Simulation
- Probabilistic Modeling including Bayesian Belief Networks
- Discrete Event Process Simulation
- Other Advanced Modeling Techniques
  - Markov, Petri-net, Neural Nets, Systems Dynamics

# Step 1 - Statistical Regression Analysis

| | | Y: Continuous | Y: Discrete |
|---|---|---|---|
| X | Discrete | ANOVA and Dummy Variable Regression | Chi-Square, Logit & Logistic Regression |
| | Continuous | Correlation & Linear Regression | Logistic Regression |

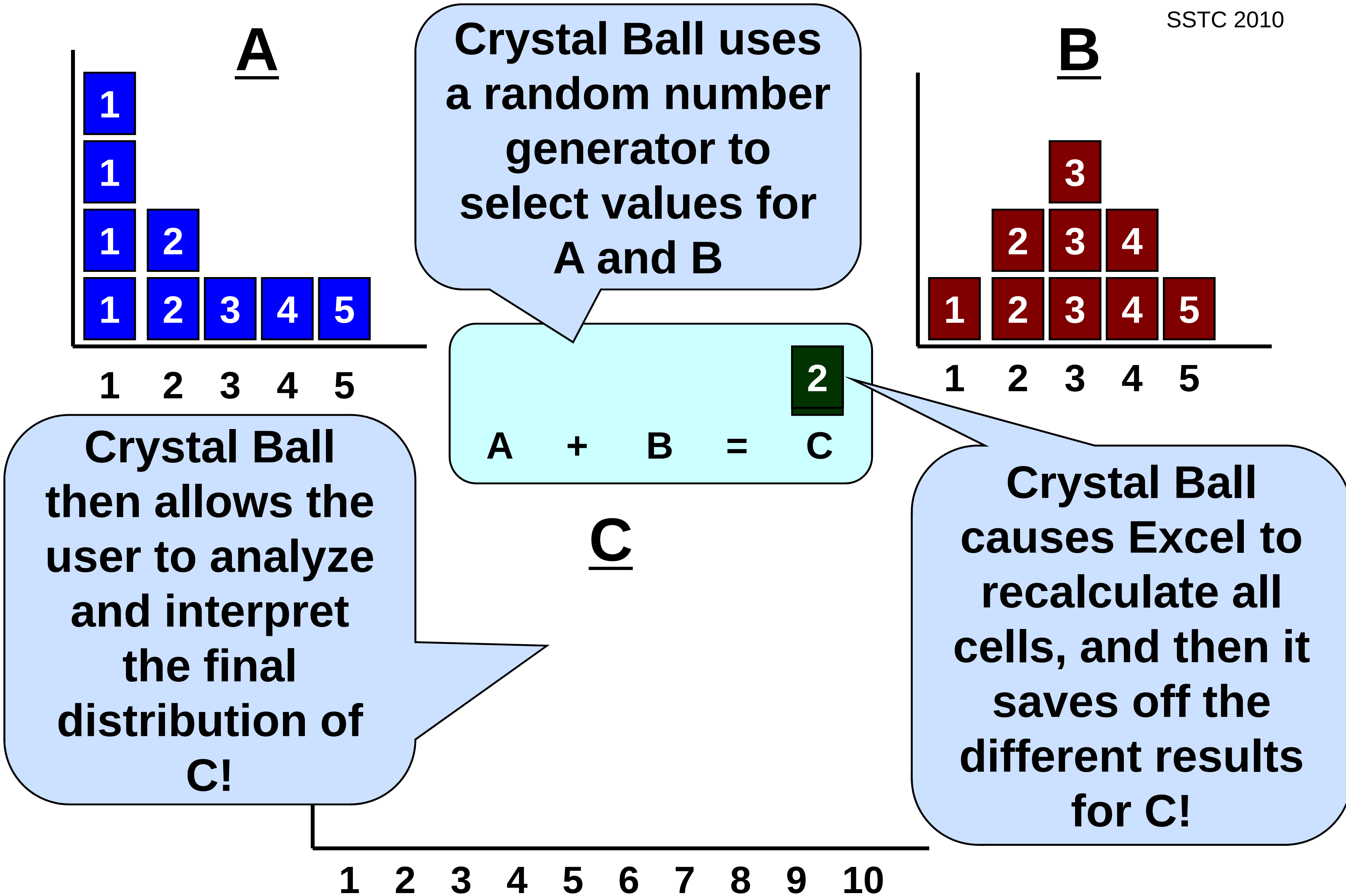


Robert Stoddard and Dave Zubrow

# Step 1 - Why Use Monte Carlo Simulation?

Use Monte Carlo simulation to do the following:

- Allow modeling of variables that are uncertain (e.g., put in a range of values instead of single value)
- Enable more accurate sensitivity analysis
- Analyze simultaneous effects of many different uncertain variables (e.g., more realistic)
- Aid buy-in and acceptance of modeling because user-provided values for uncertain variables are included in the analysis
- Provide a basis for confidence in a model output (e.g., supports risk management)
- Increase the usefulness of the model in predicting outcomes

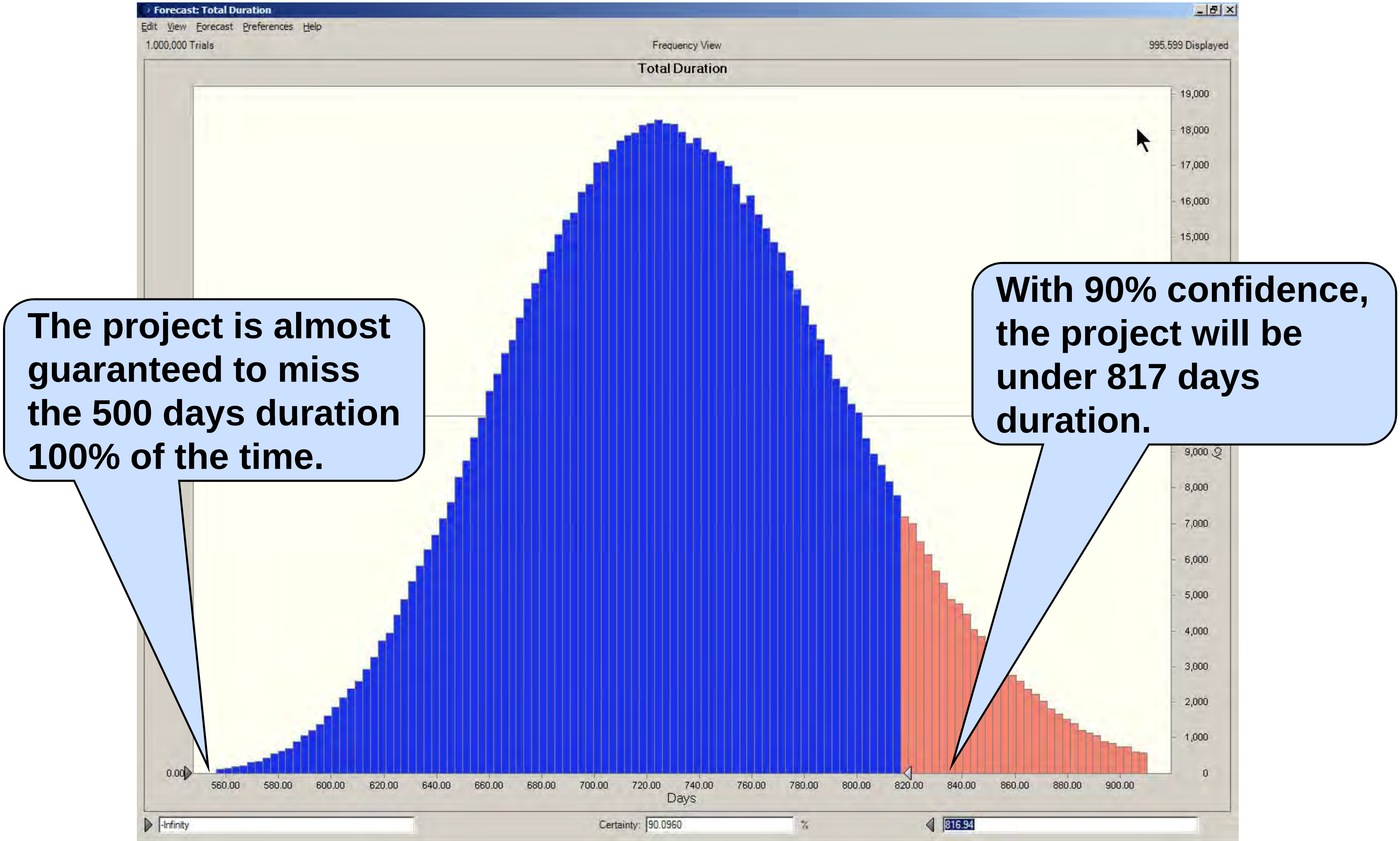

A
1
1
1 2
1 2 3 4 5
1 2 3 4 5
Crystal Ball uses a random number generator to select values for A and B
B
3
2 3 4
1 2 3 4 5
1 2 3 4 5
2
A + B = C
Crystal Ball then allows the user to analyze and interpret the final distribution of C!
C
Crystal Ball causes Excel to recalculate all cells, and then it saves off the different results for C!
1 2 3 4 5 6 7 8 9 10

# Example: Adding Reality to Schedules-1

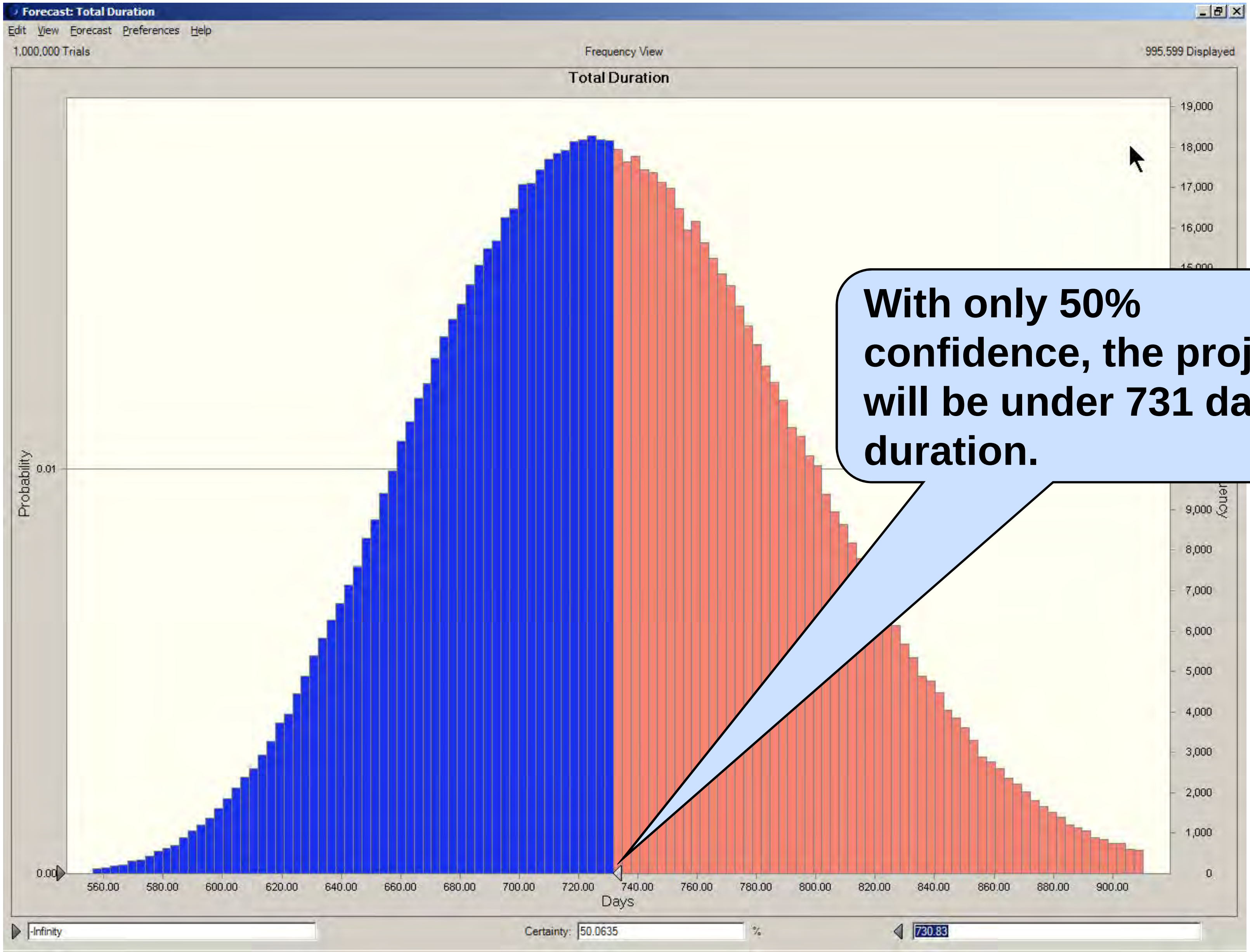
| **Process** | **Durations** | | |
|---|---|---|---|
| Step | Best | Expected | Worst |
| 1 | 27 | 30 | 75 |
| 2 | 45 | 50 | 125 |
| 3 | 72 | 80 | 200 |
| 4 | 45 | 50 | 125 |
| 5 | 81 | 90 | 225 |
| 6 | 23 | 25 | 63 |
| 7 | 32 | 35 | 88 |
| 8 | 41 | 45 | 113 |
| 9 | 63 | 70 | 175 |
| 10 | 23 | 25 | 63 |
| | | 500 | |

**What would you forecast the schedule duration to be?**

# Adding Reality to Schedules-2

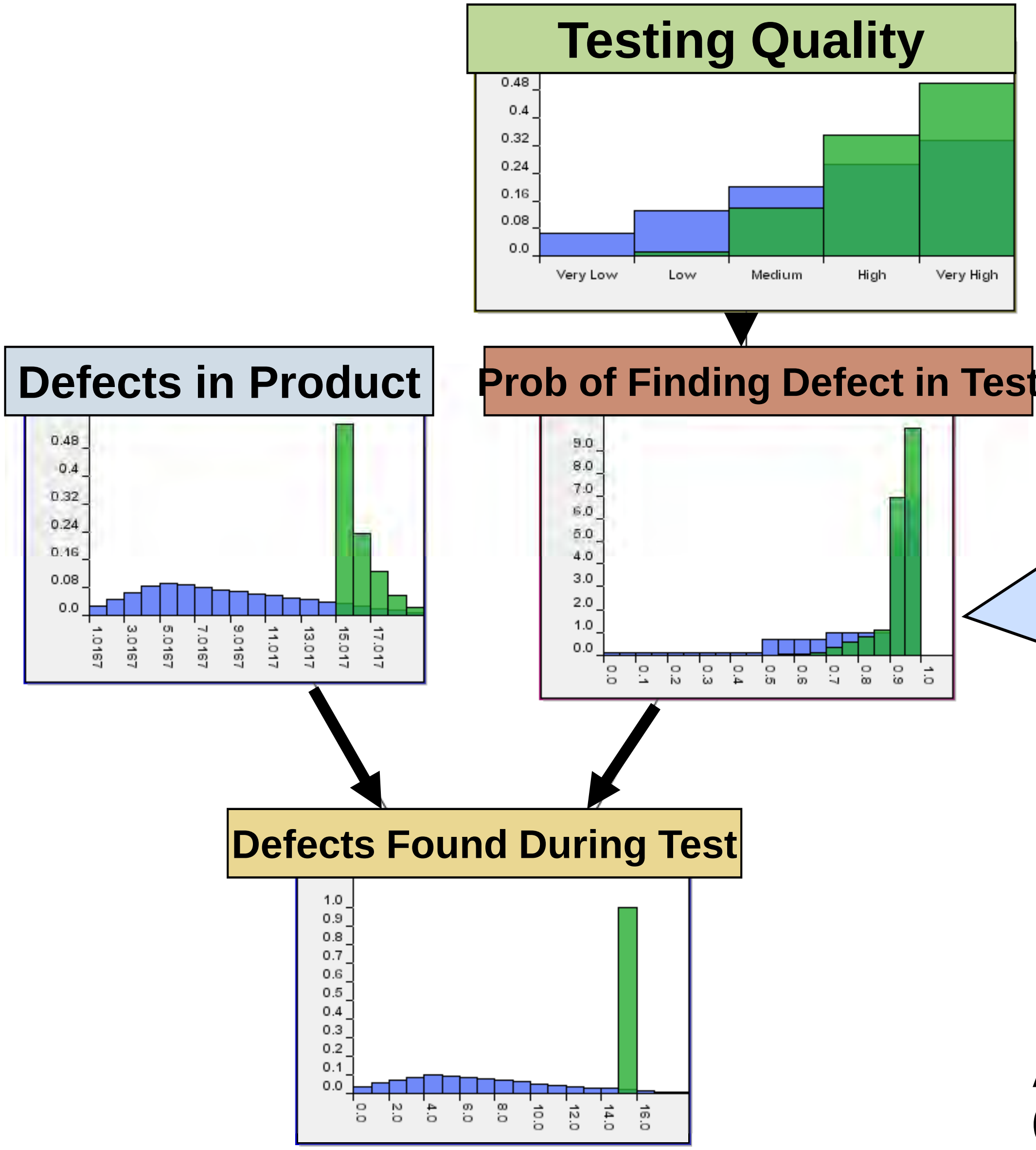


# Adding Reality to Schedules-3




With only 50% confidence, the project will be under 731 days duration.

# Example:  BBN Quality Model

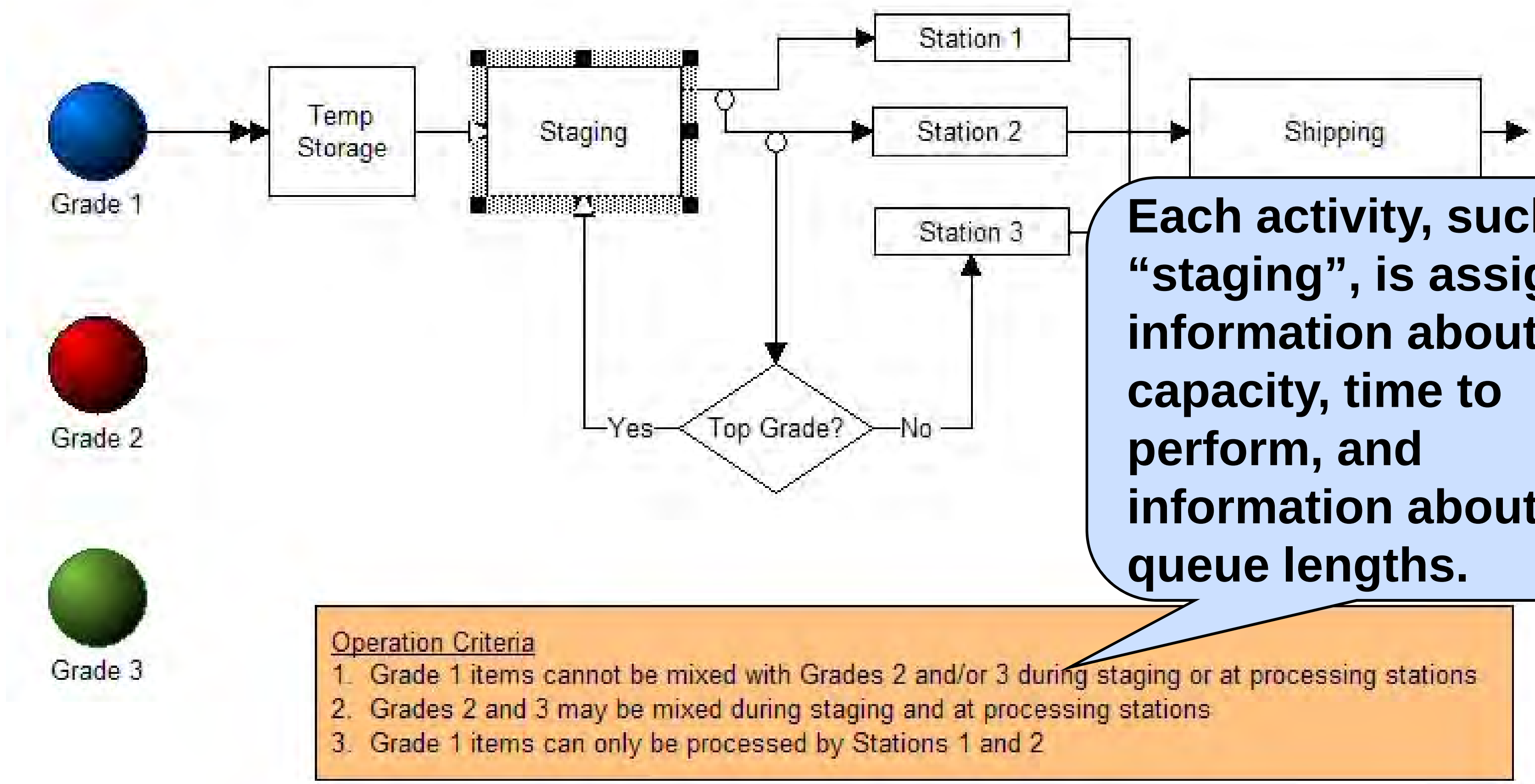



**Predict the probability of finding a defect during a test by learning what the quality of testing is.**

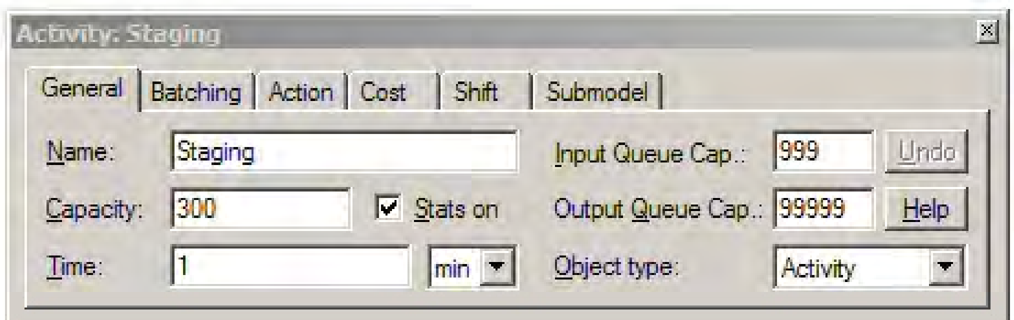
**Predict defects found by learning more about the expected incoming defect level and the ability to find defects with testing.**

**AgenaRisk. [http://www.agena.co.uk](http://www.agena.co.uk). (URL valid as of April 2007)**

# Example: Discrete Event Process Simulation




**Each activity, such as "staging", is assigned information about capacity, time to perform, and information about input queue lengths.**




**Adapted from ProcessModel, Inc. *ProcessModel*. [http://www.processmodel.com](http://www.processmodel.com). (URL valid as of April 2007)**

# Step 1 - Implement the Model in a Tool

Statistical Modeling: Example tools include Minitab, SAS JMP

Monte Carlo Simulation: Example tools include Crystal Ball and @Risk

Probabilistic Modeling: Example tools include AgenaRisk, Netica, Hugin

Discrete Event Simulation: Example tools include ProcessModel and Savvion

# Step 1 - Example Statistical Package Tools


# Step 1 - Example Monte Carlo Simulation Tools


# Step 1 - Example Probabilistic Modeling Tools

"AGENARISK" http://www.agena.co.uk/

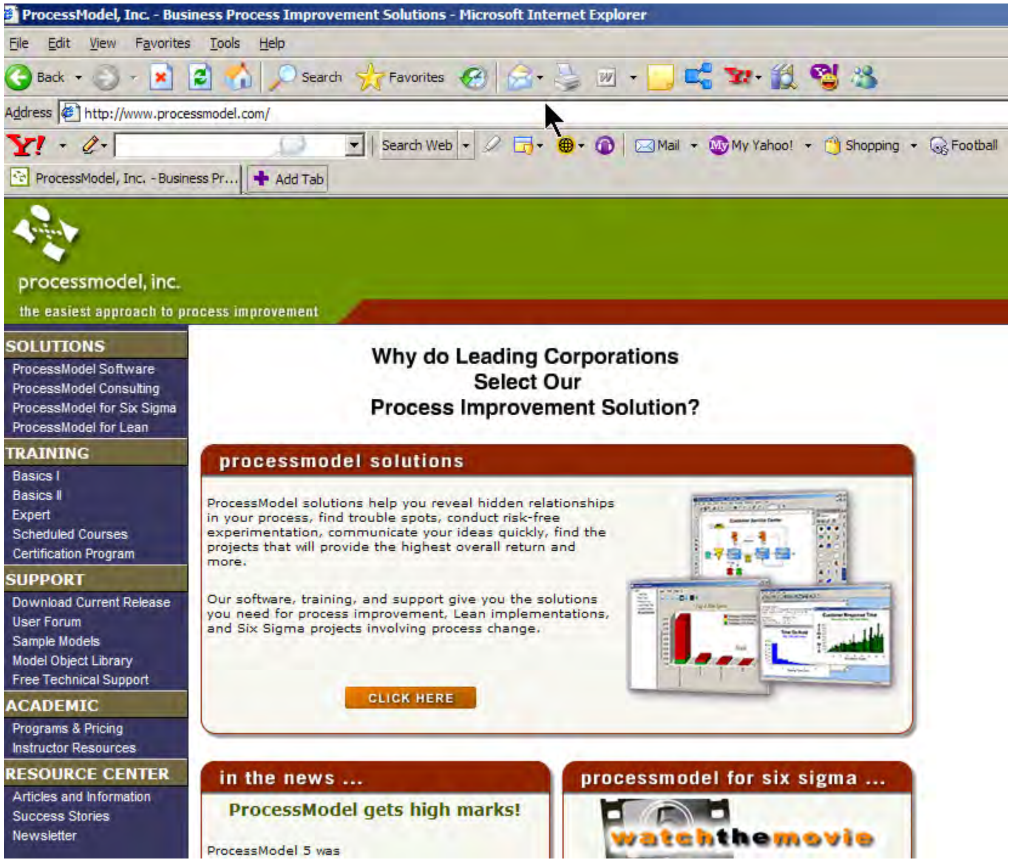

"NETICA" http://www.norsys.com/


NORSYS makes advanced Bayesian belief network and influence diagram technology practical and affordable.

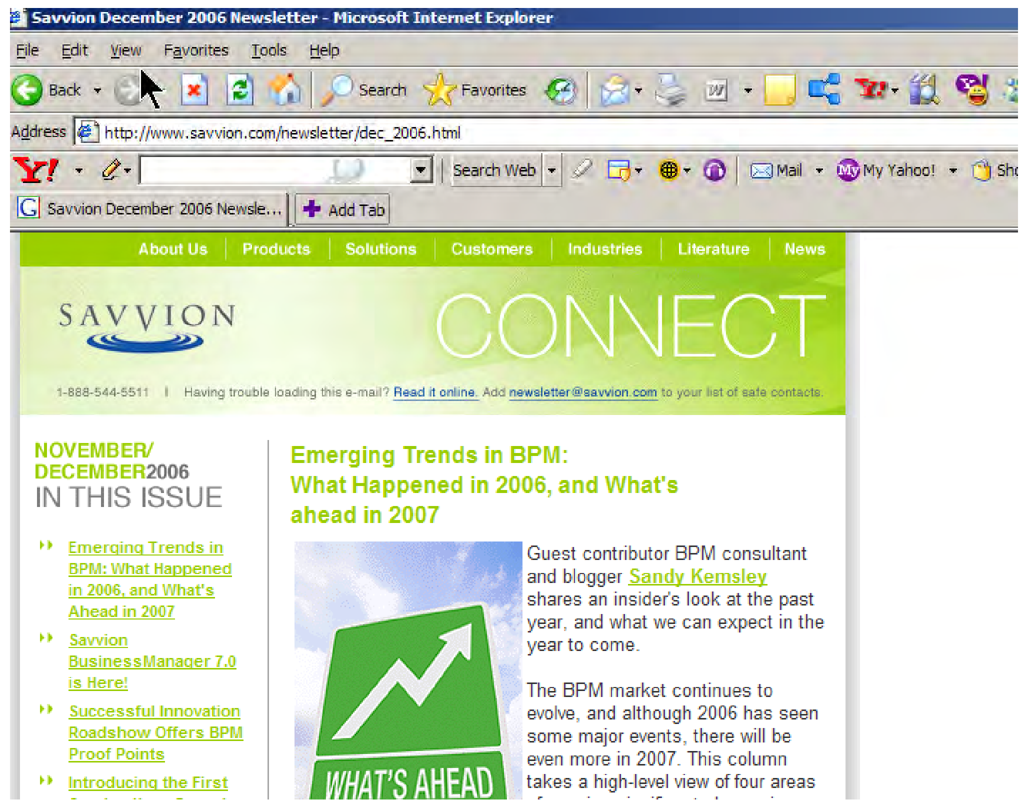
"HUGIN" http://www.hugin.com/

# Step 1 - Example Discrete Event Simulation Tools

http://www.processmodel.com




http://www.savvion.com



# Step 2 - Create Predictions with Both Confidence and Prediction Intervals-1

Because the central theme of CMMI High Maturity is understanding and controlling variation, PPMs produce statistical intervals of behavior for outcomes such that individual predicted values will have an associated confidence level

All of the Process Performance models discussed provide the ability to compute both the confidence and prediction intervals of the outcomes. These intervals are defined on the next slide

# Step 2 - Create Predictions with Both Confidence and Prediction Intervals-2

Confidence Intervals: The statistical range of behavior of a an average value computed from a sample of future data points

Prediction Intervals: The statistical range of behavior of individual future data points

**Note**: Prediction Intervals are almost always much wider than confidence intervals because averages don't experience the wide swings that individual data points can experience (similar to how individual grades in college compared to your grade point average)

# Step 3 - Validating and Maintaining PPMs - 1

Initial estimation of a PPM typically yields

- Equation or function describing the relationship between independent variables (x's) and the dependent variable (y)
- An indication of the goodness-of-fit of the model to the data (e.g., R-square, Chi-square)

These do not necessarily indicate whether the model provides sufficient practical value

- Track and compare predictions with actual results
- Failure to meet business criteria (e.g., +/- 10%) indicates need to recalibrate (i.e, same variables with different data) or remodel (new variables and data)

# Step 3 - Validating and Maintaining PPMs - 2

One strategy to jump start this process is to use half the data to estimate the model and the other half for validation (and other variations on this theme)

A second strategy is to accept that some period of time going forward will be needed to collect sample data by which to validate the PPM

# Step 4 - Confirm the PPM Meets the Healthy Ingredients

PPMs can have the greatest business benefit when they meet all of the healthy ingredients

However, PPMs should not only be evaluated in isolation, but rather, as a collection of models enabling the organization and it’s projects to most likely exhibit superior results

That said, not every PPM has to exhibit each and every healthy ingredient to be considered as a member of the portfolio of PPMs serving the organization.

# Tips - Barriers to Building PPMs

Lack of compelling outcomes to predict due to misalignment with critical business goals, usually caused by insufficient management sponsorship and involvement

Lack of a connection to a work process or sub-process such that direct changes in that process or sub-process can help cause changes in predicted outcomes

Insufficient process and domain knowledge which is necessary to identify the probable x factors to predict the outcome

Insufficient training and practice with modeling techniques

# Tips - Documentation Needed when Building PPMs-1

Similar to the existing SEI Indicator Template but with some additional information content:

1. Identity of associated processes and subprocesses
2. Identity of the outcome measure (y) and the x factors
3. Data type of all outcome (y) and x factors
4. Statistical evidence that the x factors are significant (e.g. p values of individual x factors)
5. Statistical evidence of the strength of the model (e.g. the adjusted R-squared value)
6. The actual prediction equation for the outcome (y)
7. The performance baselines of the x factors



Robert Stoddard and Dave Zubrow

# Tips - Documentation Needed when Building PPMs-2

Similar to the existing SEI Indicator Template but with some additional information content (continued):

8. The resulting confidence interval of the predicted outcome
9. The resulting prediction interval of the predicted outcome
10. Use case scenarios of how the PPM is intended to be used by different audiences for specific decisions
11. Description of how often the PPM is updated, validated, and calibrated
12. Description of how often the PPM is used to make predictions with results shown to decision-makers
13. Description of which organizational segment of projects the PPM applies to

# Overview of the Steps to Build PPMs

# - Using PPMs

Software Engineering Institute | Carnegie Mellon

# Using PPMs

Use these models to assist with statistical management of critical subprocesses

Use the predictions of these models to make decisions and take preventive and mitigative action

Use these models to help with CAR and OID

Coach audiences on how to understand, interpret and draw conclusions from process performance models

# Take Action Based on Results of PPM Predictions

If a PPM model predicts an unacceptable range of values for a particular outcome, then early action can influence a more desirable range of outcome

Once a PPM model predicts a range of values for a particular outcome, then actual values can be compared to the range.   If the actual values fall outside the range, it may be treated similarly to a point on a control chart falling outside of the control limits

Use PPM predictions to help inform process composition decisions so that business goals may be optimized

# How PPMs Assist CAR

- Aid impact, benefit, and ROI predictions for
  - Selecting defects for analysis
  - Selecting action proposals for implementation
- Use PPMs to identify potential sources of the problem or defect
- Use PPMs to understand the interactions among selected improvements; and the combined predicted impacts, costs, and benefits of the improvements (considered as a set)
- Compare the result versus the original PPM-based prediction

# How PPMs Assist OID

- Select process improvement proposals for implementation by aiding impact, benefit, and ROI predictions
- Identify opportunities for improvement
- Use PPMs to understand the interactions among selected improvements; and the combined predicted impacts, costs, and benefits of the improvements (considered as a set)
- Prioritize improvements based on ROI, cost, risk, etc.
- Confirm the prediction (provides input to maintaining PPMs)

# What is Sub-optimization and how can PPMs help?

Sub-optimization is where one parameter is optimized at the expense of other(s)

- Reduce delivered defects, but are late and over budget
- Meet the cost goal but don’t deliver desired functionality

PPMs allow you to

- Gage the trade-offs amongst multiple goals
- Gage the effects of changes to multiple parameters

# PPM Exercise 1: Constructing a Product Business Case with Monte Carlo Simulation and Optimization

# Monte Carlo Simulation Steps with Crystal Ball

Start

(1) Start Crystal Ball

(2) Define a simulation model
(2.1) Define assumption cells
(2.2) Select subprocess options
(2.3) Define forecast cells

(3) Run simulations
(3.1) Set run preferences
(3.2) Run simulations
(3.3) Save & restore simulation results

(4) Analyze simulation results
(4.1) Understand and use forecast charts
(4.2) Determine the certainty level
(4.3) Create reports
(4.4) Review simulation results

End

At this time, launch Crystal Ball which will automatically launch Excel and then add itself in.

# Crystal Ball Toolbar



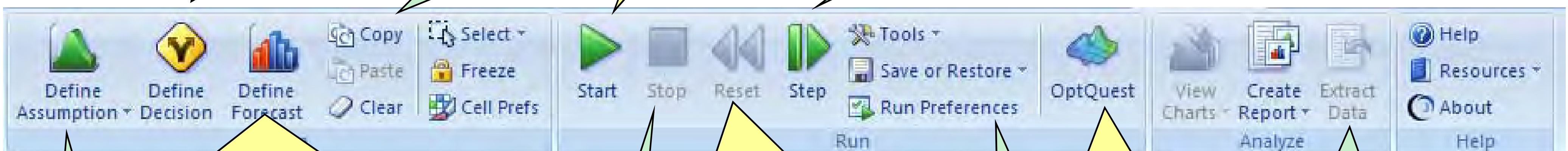
<u>Define decision</u>
(Lets you identify a cell as a decision cell to be used in Optimization Modeling)

You can <u>copy, paste and clear</u> Crystal Ball identities to save time

<u>Start simulation</u>
(Start simulation once all settings are made)

<u>Single step</u>
(Lets you run the simulation step by step. Normally used to debug issues with simulation)

<u>Create report</u>
(Creates standardized reports of the simulation)




<u>Define forecast</u>
(Lets you identify a cell as an outcome that you want to study)

<u>Reset simulation</u>
(Restart simulation and erase previous results)

<u>Optquest</u>
(Begin optimization)

<u>Extract data</u>
(Allows the capture and saving of the actual simulation data from all the runs)

<u>Define assumption</u>
(Lets you identify a cell as an uncertain cell with a distribution)

<u>Stop simulation</u>
(You can stop the simulation midstream)

<u>Run preferences</u>
(Enables the settings of how long the simulation runs, etc…)

# Benefits of Using Optimization Modeling

Monte Carlo simulation models can only provide a range of possible outcomes for any situation. They do not identify ways to **control** the situation to achieve the **best** outcome.

Optimization modeling

- automates tens of thousands of decision “what-ifs” from a Monte Carlo simulation to determine the best possible solution
- is easy to use, not tedious and time consuming like many other analytical methods
- uses state-of-the-art algorithms for confidently finding optimal solutions
- supports decision making in situations where significant resources, costs, or revenues are at stake

# Steps for Optimization Using Crystal Ball



(1) Create a simulation model of the problem.

↓

(2) Define decision variables cells.

↓

(3) Select the objective for the optimization.

↓

(4) Identify additional requirements.

↓

(5) Confirm settings for decision variables.

↓

(6) Specify constraints for decision variables.

↓

(7) Identify Optimization Parameters.

↓

(8) Run the Optimization.

↓

(9) Interpret the Results.

## Business Case Example for Feature Inclusion Decision in Upcoming Hospital Records Software Project

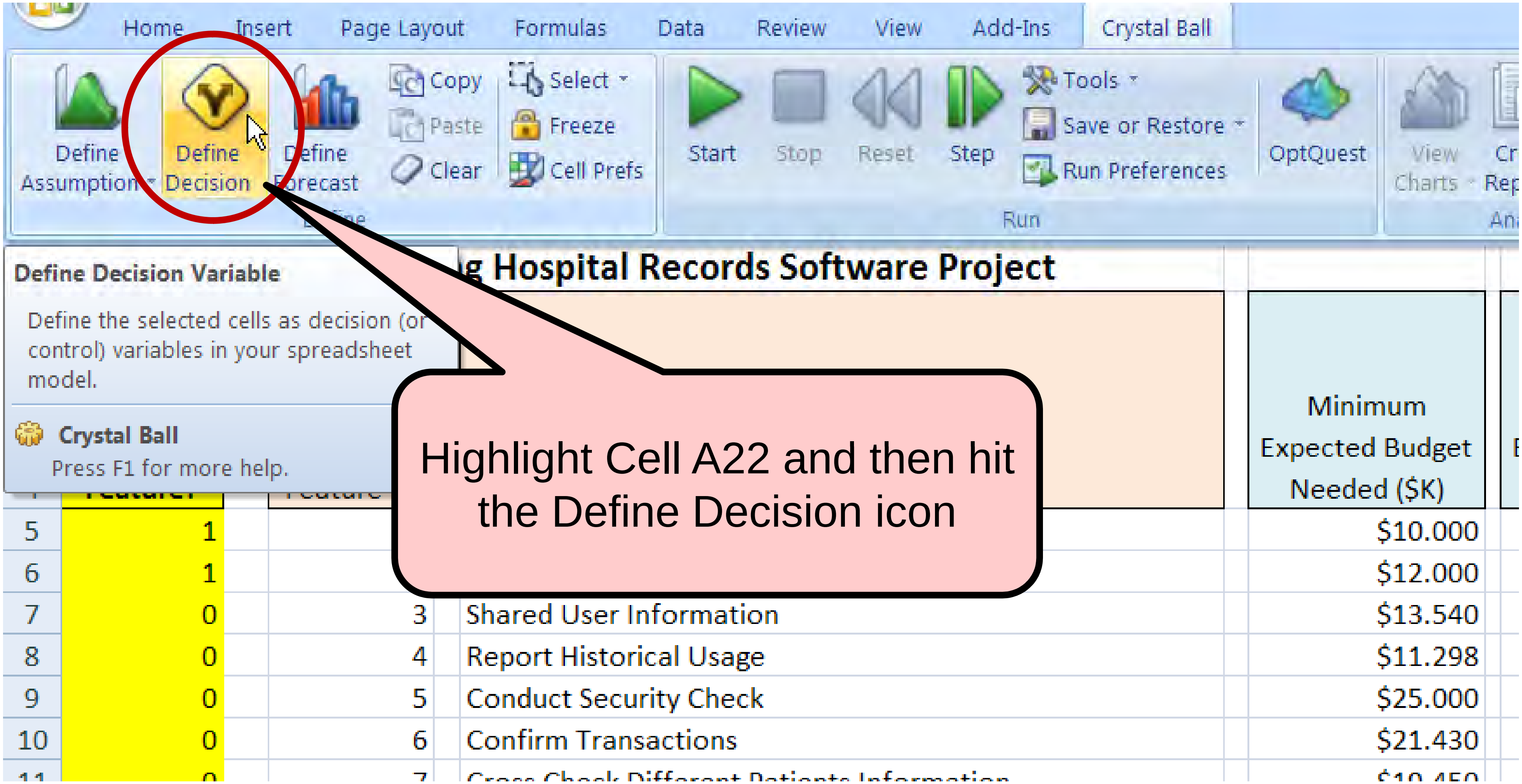
**Business Case Monte Carlo Optimization-v010.xls file**

| | Develop Feature? | Feature ID | Feature Description |
|---|---|---|---|
| 5 | 1 | 1 | Online Web Access |
| 6 | 1 | 2 | Real-time Updating of Information |
| 7 | 0 | 3 | Shared User Information |
| 8 | 0 | 4 | Report Historical Usage |
| 9 | 0 | 5 | Conduct Security Check |
| 10 | 0 | 6 | Confirm Transactions |
| 11 | 0 | 7 | Cross Check Different Patients Information |
| 12 | 0 | 8 | Trace Prescriptions Used |
| 13 | 0 | 9 | Trace Assigned Doctor |
| 14 | 0 | 10 | Trace Hospital |
| 15 | 1 | 11 | Conduct Periodic Audit |
| 16 | 1 | 12 | Check for Corrupt Data |
| 17 | 0 | 13 | Provide Conflict Warning |
| 18 | 0 | 14 | Identify Incomplete Records |
| 19 | 1 | 15 | Compute Cycle Times on Value Stream |
| 20 | 1 | 16 | Enable cross hospital sharing of data |
| 21 | 0 | 17 | Provide Security Encryption for Sensitive Data |
| 22 | 1 | 18 | Enable workflow automation messages |
| 23 | 1 | 19 | Require peer review of critical data inputs |
| 24 | 1 | 20 | Provide for automated archival of information |

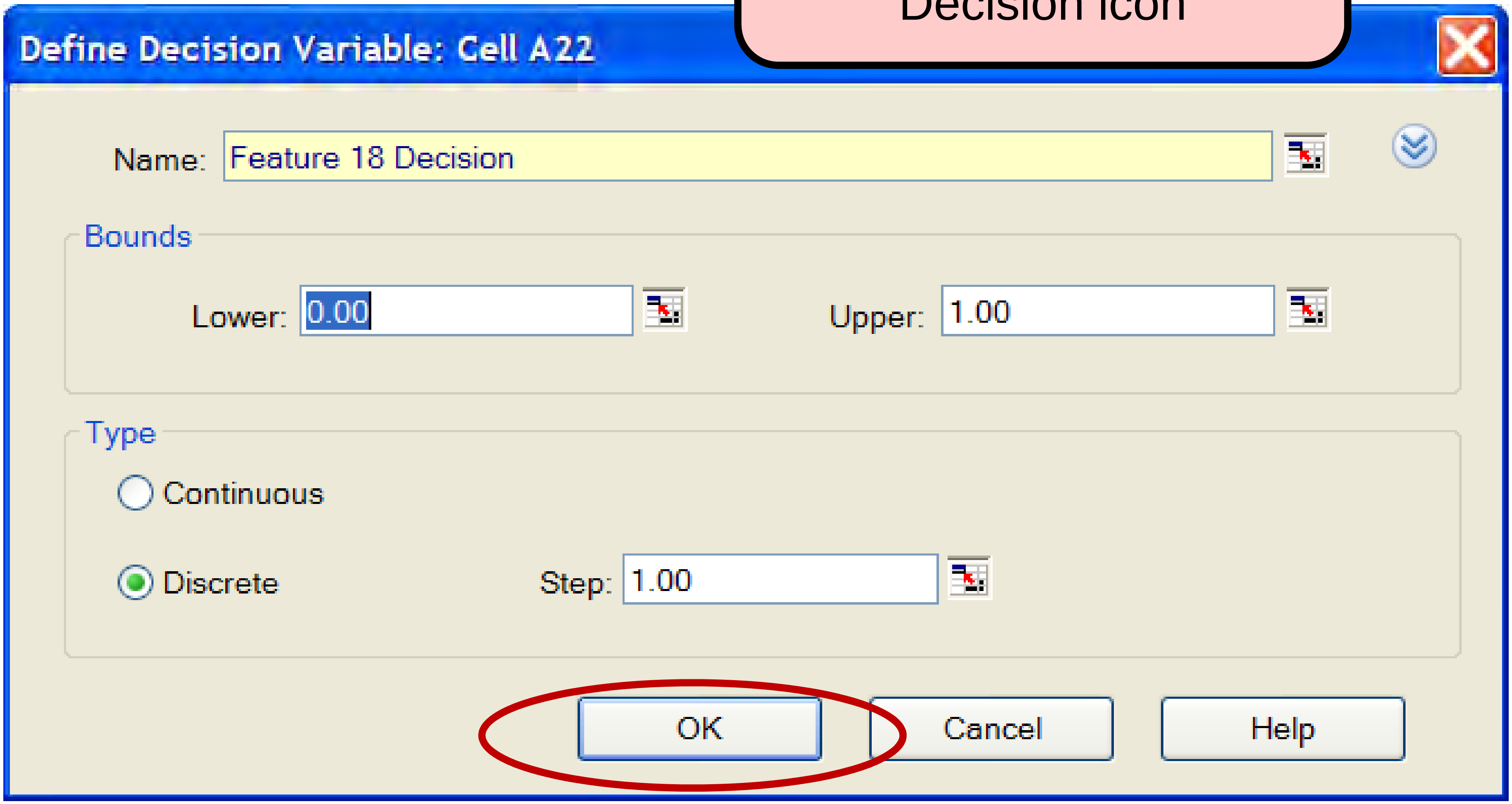
| Minimum Expected Budget Needed ($K) | Most Likely Expected Budget Needed ($K) | Maximum Expected Budget Needed ($K) | **Simulated Budget ($K)** | **Budget Used in Simulation Feature Scenario ($K)** |
|---|---|---|---|---|
| $10.000 | $12.000 | $29.260 | $0.000 | $0.000 |
| $12.000 | $14.400 | $21.870 | $0.000 | $0.000 |
| $13.540 | $16.248 | $27.420 | $0.000 | $0.000 |
| $11.298 | $13.558 | $19.880 | $0.000 | $0.000 |
| $25.000 | $30.000 | $35.290 | $0.000 | $0.000 |
| $21.430 | $25.716 | $29.830 | $0.000 | $0.000 |
| $19.450 | $23.340 | $39.750 | $0.000 | $0.000 |
| $18.390 | $22.068 | $38.234 | $0.000 | $0.000 |
| $17.420 | $20.904 | $29.774 | $0.000 | $0.000 |
| $29.170 | $35.004 | $51.960 | $0.000 | $0.000 |
| $26.290 | $31.548 | $62.948 | $0.000 | $0.000 |
| $21.290 | $25.548 | $39.497 | $0.000 | $0.000 |
| $21.990 | $26.388 | $34.659 | $0.000 | $0.000 |
| $27.990 | $33.588 | $39.774 | $0.000 | $0.000 |
| $39.230 | $47.076 | $57.849 | $0.000 | $0.000 |
| $41.090 | $49.308 | $72.895 | $0.000 | $0.000 |
| $38.210 | $45.852 | $67.391 | $0.000 | $0.000 |
| $31.280 | $37.536 | $47.324 | $0.000 | $0.000 |
| $31.670 | $38.004 | $49.846 | $0.000 | $0.000 |
| $27.720 | $33.264 | $39.888 | $0.000 | $0.000 |
| | $581.350 | | | |
| | | **Total Budget >>>>>>>>>>>>>>>>>>** | | **$0.000** |

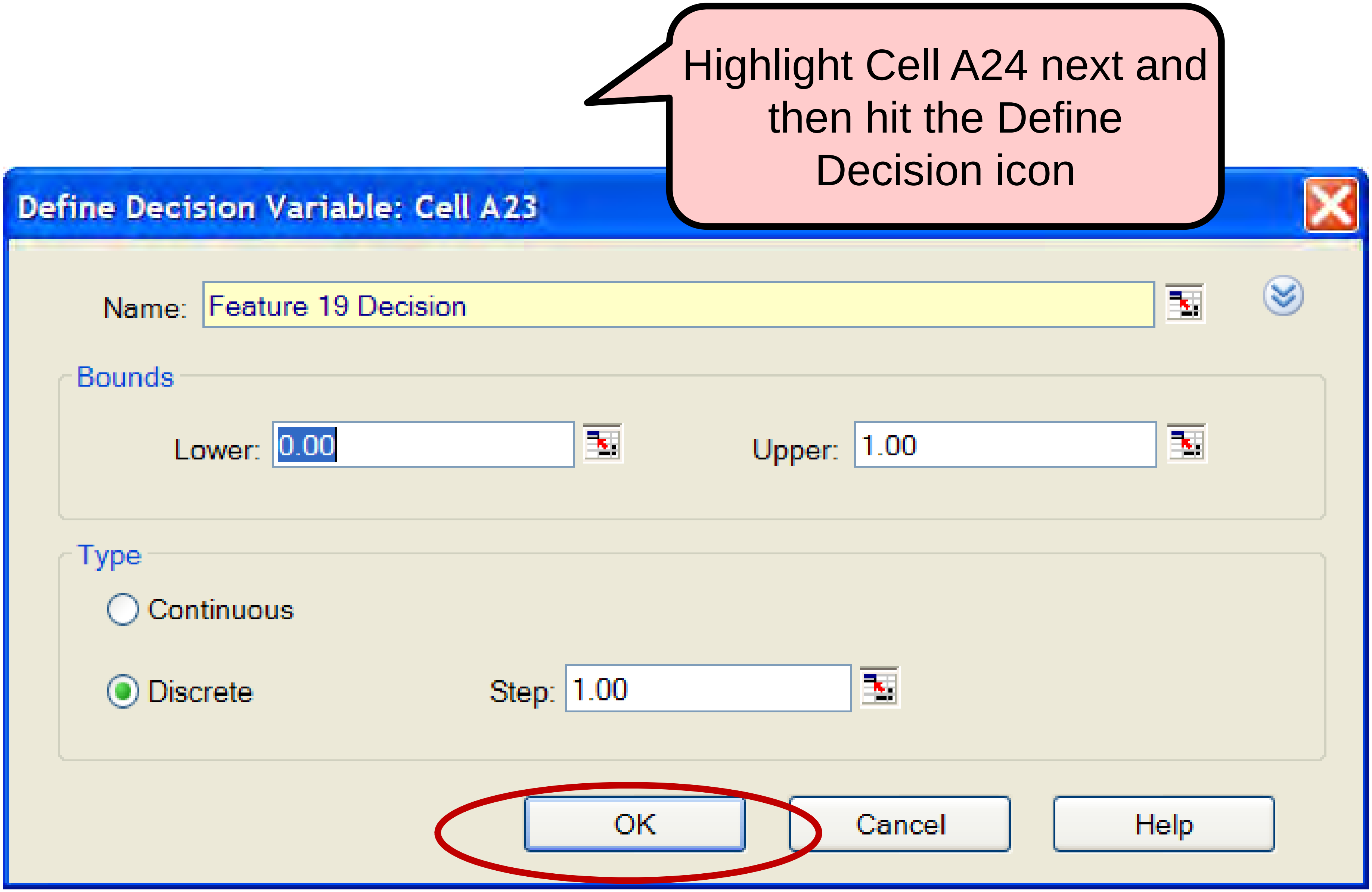
| Minimum Expected Calendar Days Needed | Most Likely Expected Calendar Days Needed | Maximum Expected Calendar Days Needed | **Simulated Calendar Days** | **Calendar Days used in Simulation Feature Scenario** |
|---|---|---|---|---|
| 15 | 20 | 30 | 0 | 0 |
| 10 | 18 | 36 | 0 | 0 |
| 12 | 15 | 56 | 0 | 0 |
| 25 | 38 | 49 | 0 | 0 |
| 30 | 38 | 75 | 0 | 0 |
| 26 | 30 | 48 | 0 | 0 |
| 18 | 29 | 62 | 0 | 0 |
| 15 | 22 | 36 | 0 | 0 |
| 19 | 26 | 39 | 0 | 0 |
| 7 | 14 | 31 | 0 | 0 |
| 28 | 37 | 45 | 0 | 0 |
| 22 | 29 | 49 | 0 | 0 |
| 26 | 40 | 67 | 0 | 0 |
| 33 | 40 | 78 | 0 | 0 |
| 18 | 26 | 40 | 0 | 0 |
| 17 | 21 | 38 | 0 | 0 |
| 26 | 29 | 37 | 0 | 0 |
| 21 | 35 | 59 | 0 | 0 |
| 22 | 29 | 51 | 0 | 0 |
| 26 | 30 | 47 | 0 | 0 |
| | | | | |
| | | **Total Calendar Days >>>>>>>>>>>>>>** | | **0** |

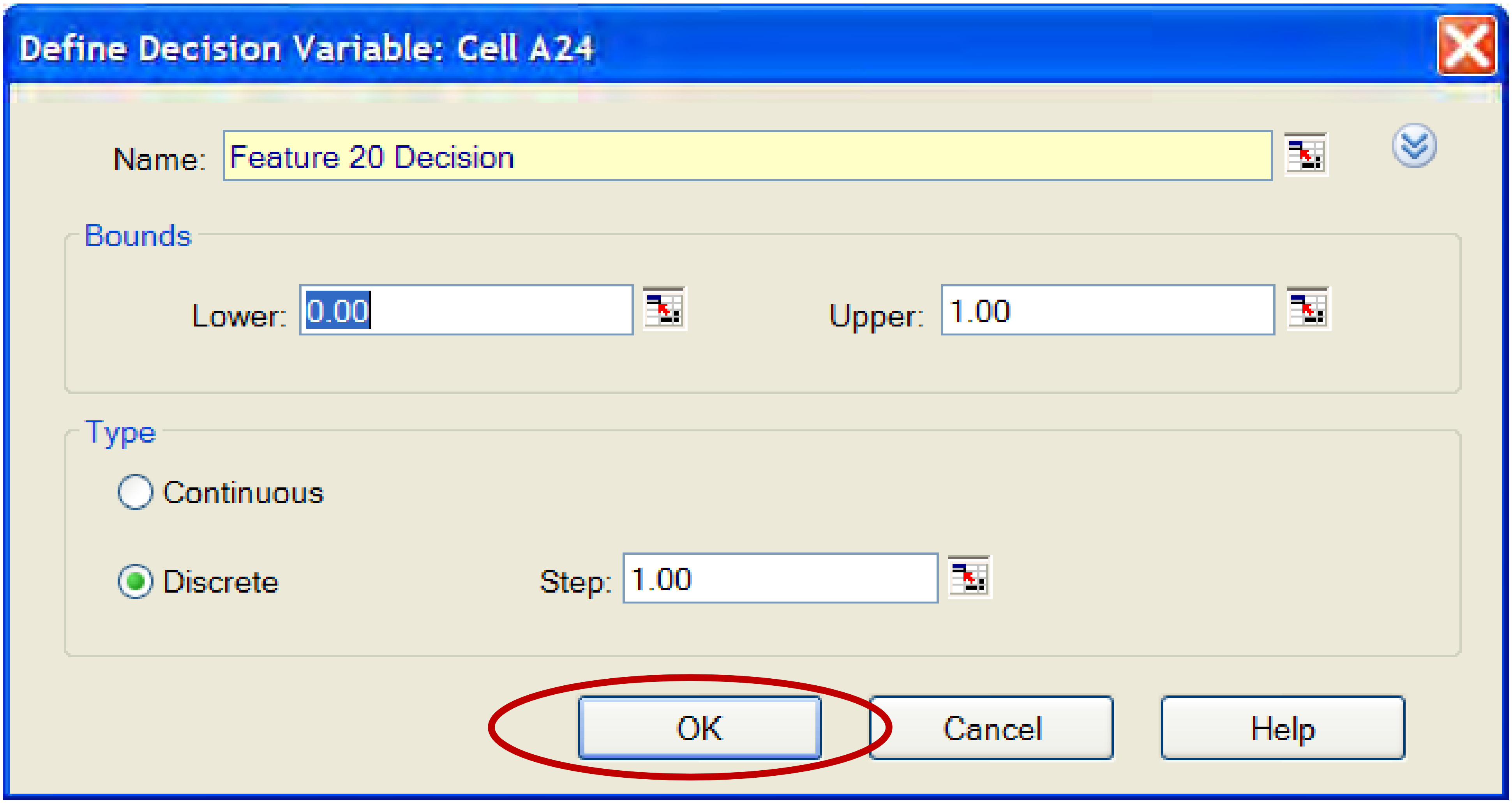
| Expected Senior Resource Needed | Actual Senior Resource Used in Simulation | Relative Customer Value | Customer Value in Simulation |
|---|---|---|---|
| 0 | 0 | 1.00 | 1.00 |
| 0.2 | 0.2 | 2.00 | 2.00 |
| 0.3 | 0 | 1.20 | 0.00 |
| 0 | 0 | 1.50 | 0.00 |
| 0.12 | 0 | 1.80 | 0.00 |
| 0.15 | 0 | 0.90 | 0.00 |
| 0.19 | 0 | 0.30 | 0.00 |
| 0.25 | 0 | 0.80 | 0.00 |
| 0 | 0 | 1.70 | 0.00 |
| 0 | 0 | 1.20 | 0.00 |
| 0 | 0 | 1.90 | 1.90 |
| 0.65 | 0.65 | 2.40 | 2.40 |
| 0.34 | 0 | 2.70 | 0.00 |
| 0.29 | 0 | 3.00 | 0.00 |
| 0.21 | 0.21 | 2.20 | 2.20 |
| 0.17 | 0.17 | 1.70 | 1.70 |
| 0 | 0 | 1.95 | 0.00 |
| 0 | 0 | 2.67 | 2.67 |
| 0 | 0 | 4.00 | 4.00 |
| 0 | 0 | 2.39 | 2.39 |
| | | | |
| **Total Resource>>** | **1.23** | **Value>>>** | **20.26** |

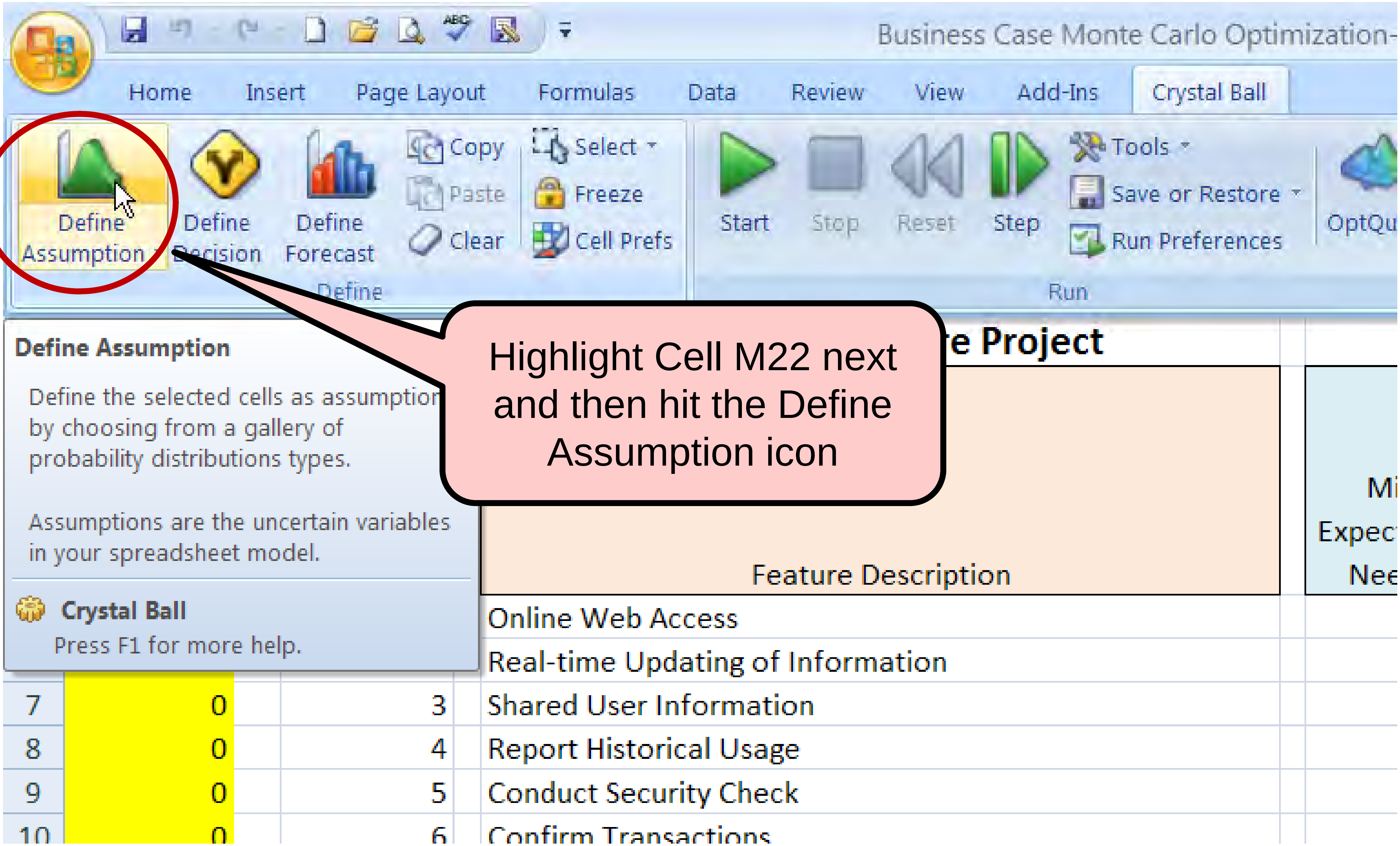

Home
Insert
Page Layout
Formulas
Data
Review
View
Add-Ins
Crystal Ball
Define Assumption
Define Decision
Define Forecast
Copy
Paste
Clear
Select
Freeze
Cell Prefs
Start
Stop
Reset
Step
Tools
Save or Restore
Run Preferences
Run
OptQuest
View Charts
Define Decision Variable
Define the selected cells as decision (or control) variables in your spreadsheet model.
Crystal Ball
Press F1 for more help.
g Hospital Records Software Project
Minimum Expected Budget Needed ($K)
Highlight Cell A22 and then hit the Define Decision icon
$10.000
$12.000
Shared User Information
$13.540
Report Historical Usage
$11.298
Conduct Security Check
$25.000
Confirm Transactions
$21.430

Highlight Cell A23 next and then hit the Define Decision icon

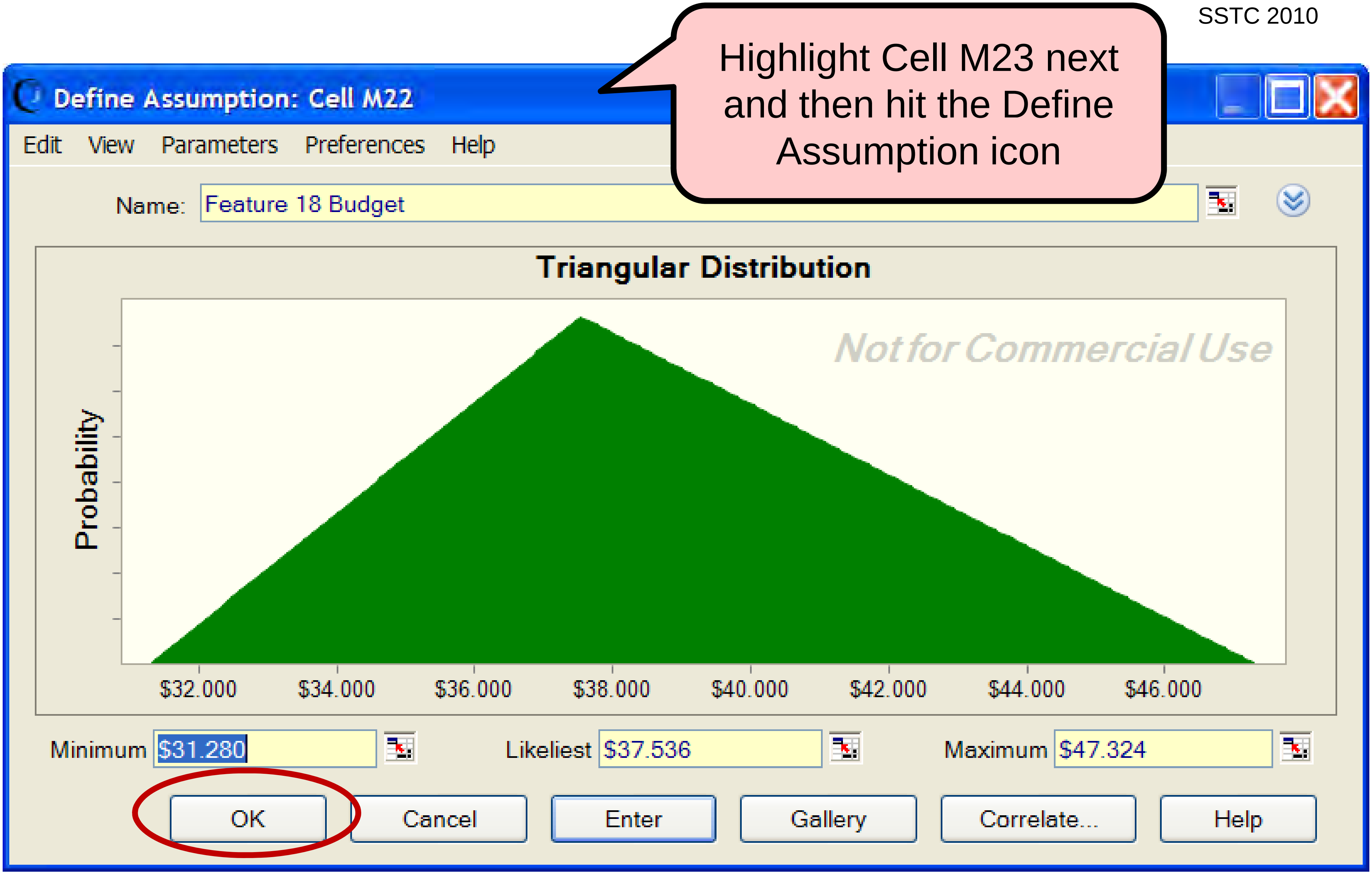


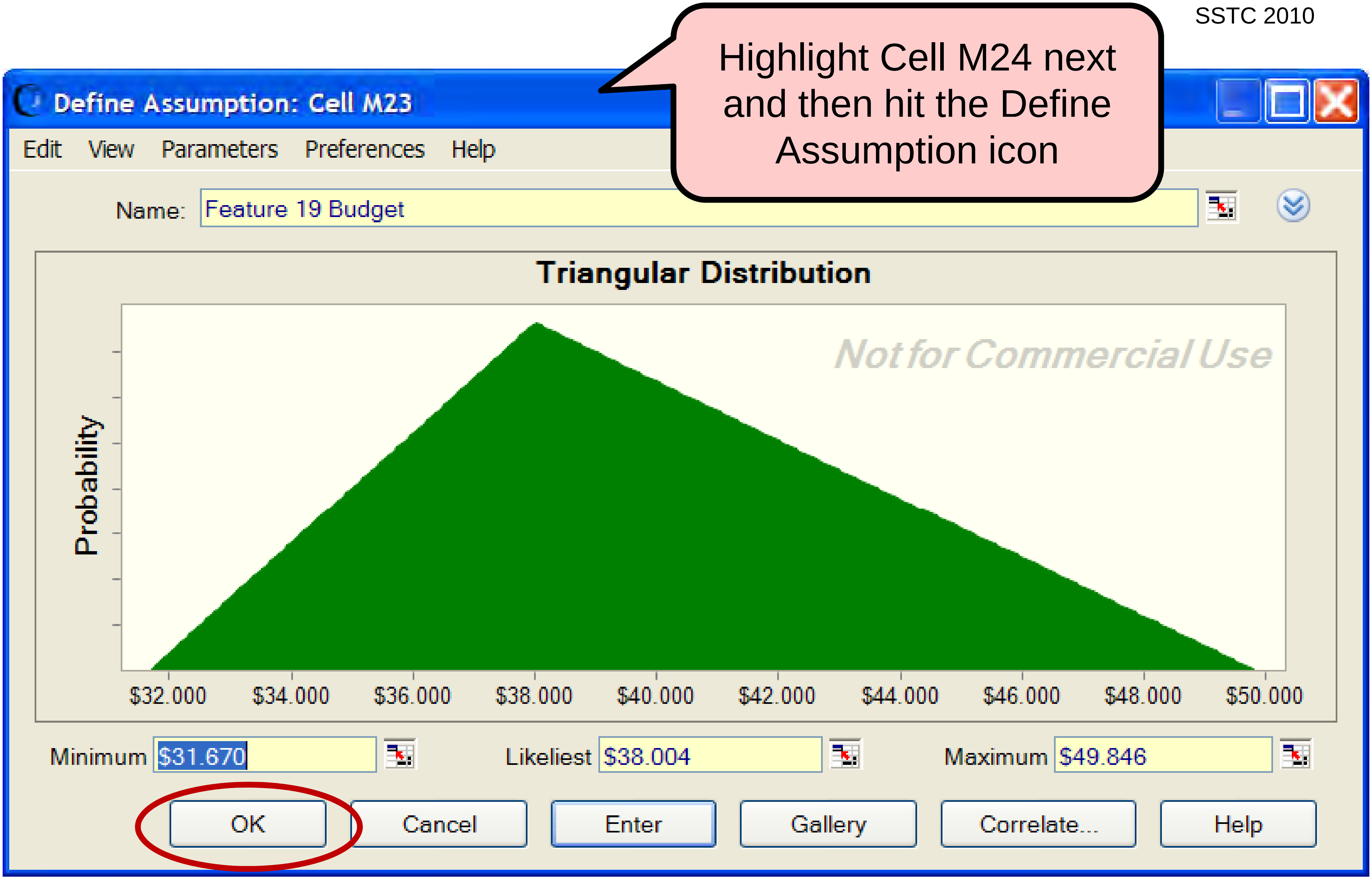

Highlight Cell A24 next and then hit the Define Decision icon
Define Decision Variable: Cell A23
Name:
Feature 19 Decision
Bounds
Lower:
0.00
Upper:
1.00
Type
Continuous
Discrete
Step:
1.00
OK
Cancel
Help

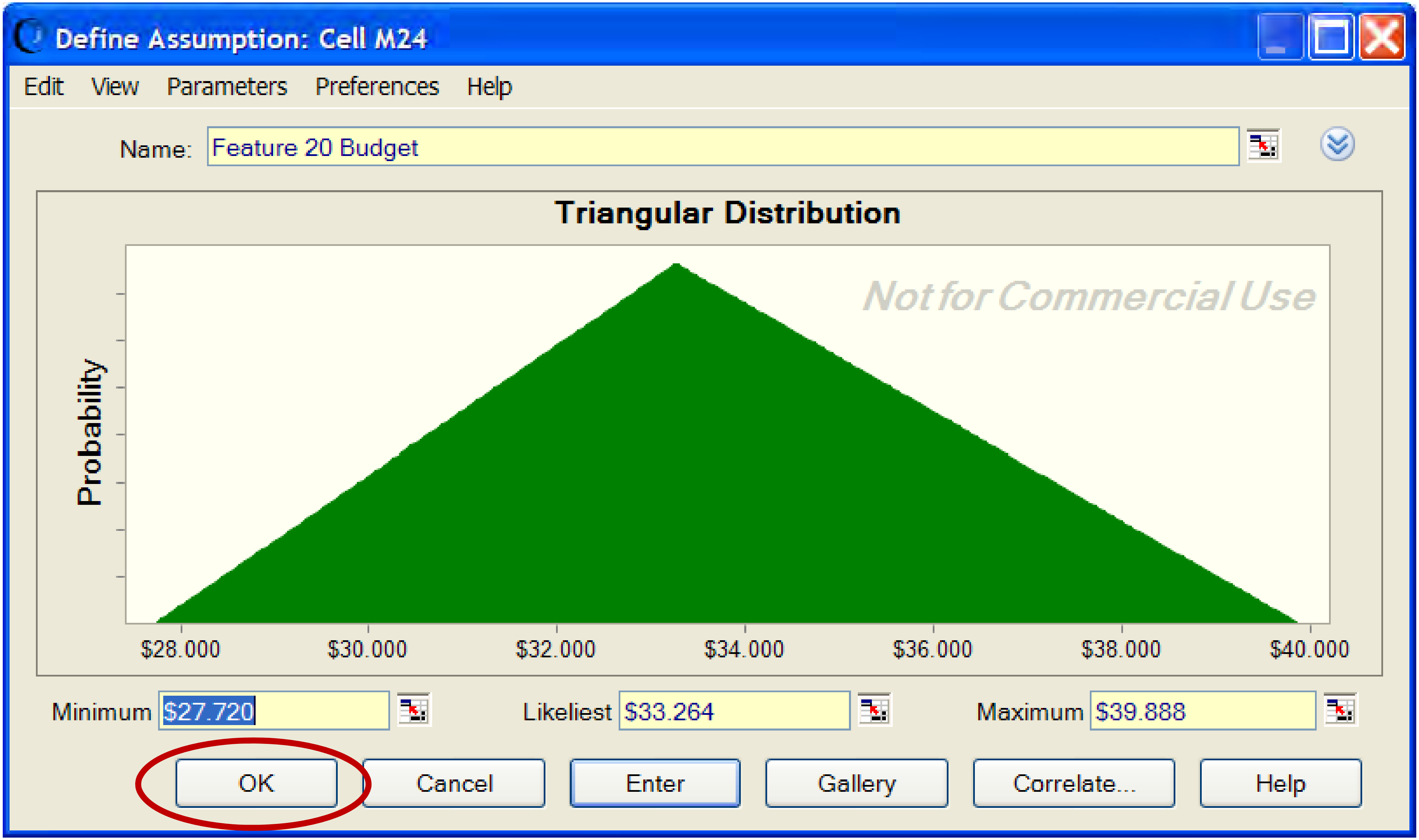

Define Decision Variable: Cell A24
Name:
Feature 20 Decision
Bounds
Lower:
0.00
Upper:
1.00
Type
Continuous
Discrete
Step:
1.00
OK
Cancel
Help

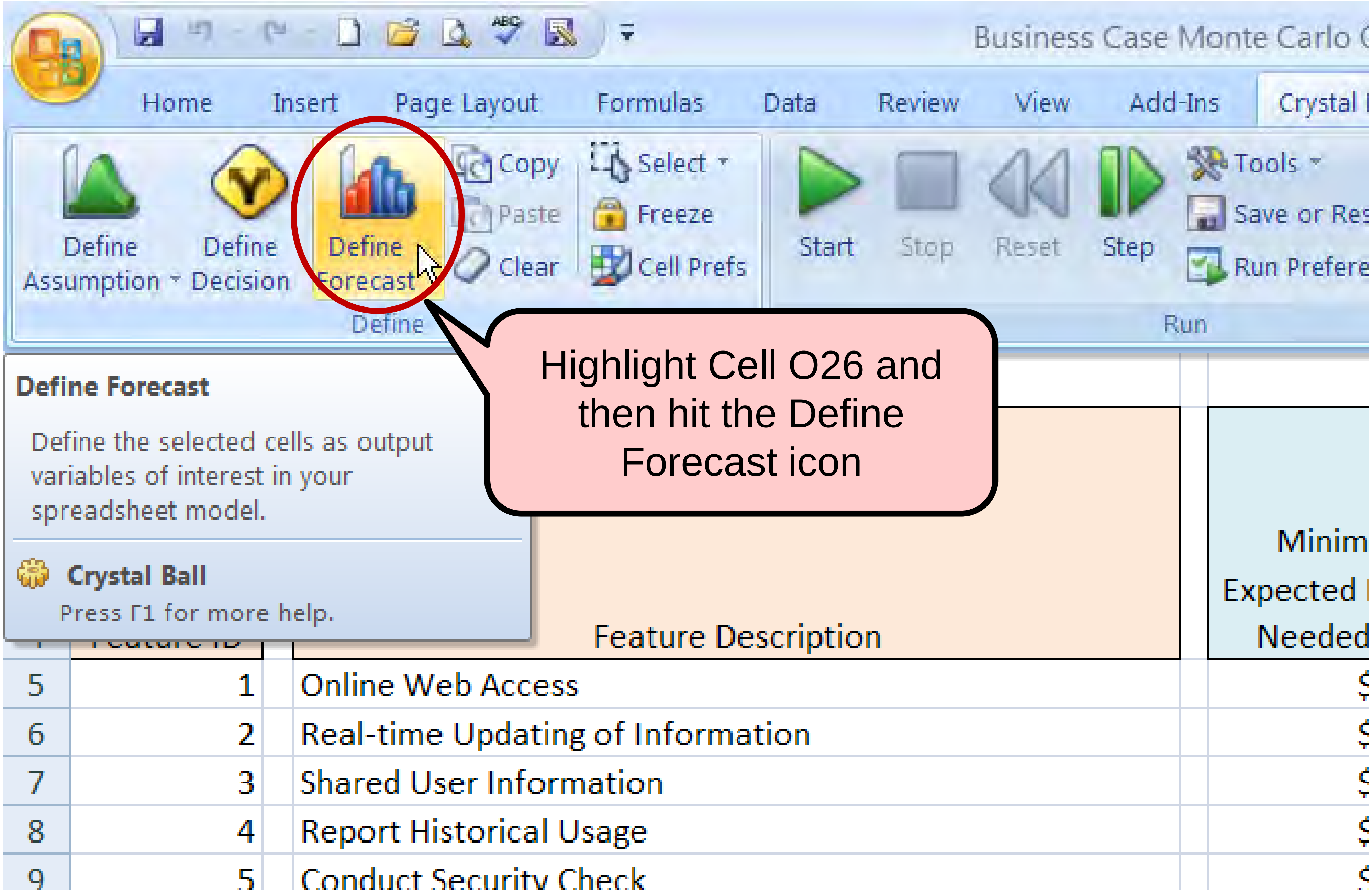



Highlight Cell M22 next and then hit the Define Assumption icon

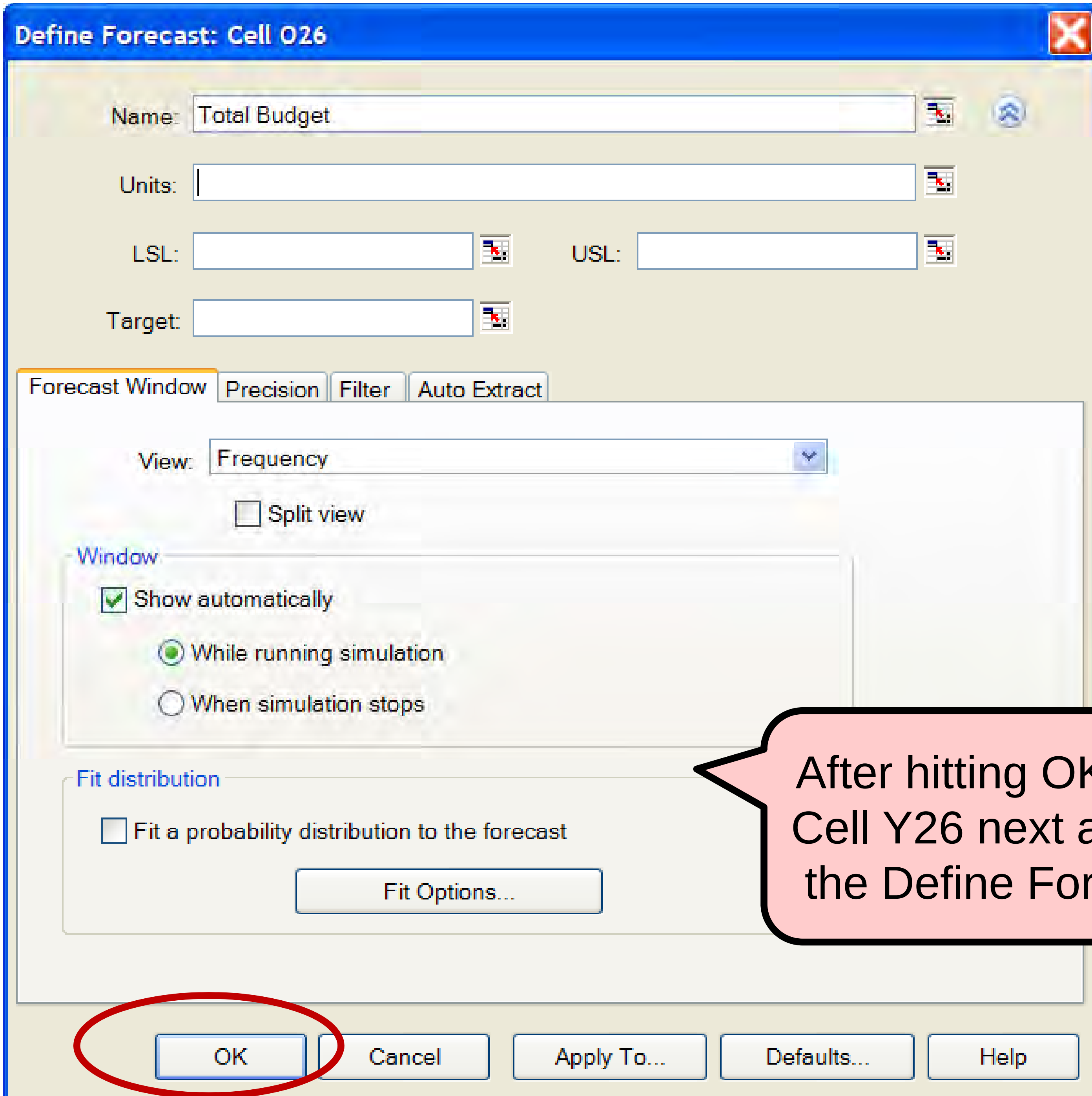

Define Assumption: Cell M22
Edit
View
Parameters
Preferences
Help
Name:
Feature 18 Budget
Triangular Distribution
Not for Commercial Use
Probability
$32.000
$34.000
$36.000
$38.000
$40.000
$42.000
$44.000
$46.000
Minimum
$31.280
Likeliest
$37.536
Maximum
$47.324
OK
Cancel
Enter
Gallery
Correlate...
Help
Highlight Cell M23 next and then hit the Define Assumption icon


Define Assumption: Cell M23
Highlight Cell M24 next and then hit the Define Assumption icon
Edit View Parameters Preferences Help
Name: Feature 19 Budget
Triangular Distribution
Not for Commercial Use
Probability
$32.000 $34.000 $36.000 $38.000 $40.000 $42.000 $44.000 $46.000 $48.000 $50.000
Minimum $31.670
Likeliest $38.004
Maximum $49.846
OK Cancel Enter Gallery Correlate... Help

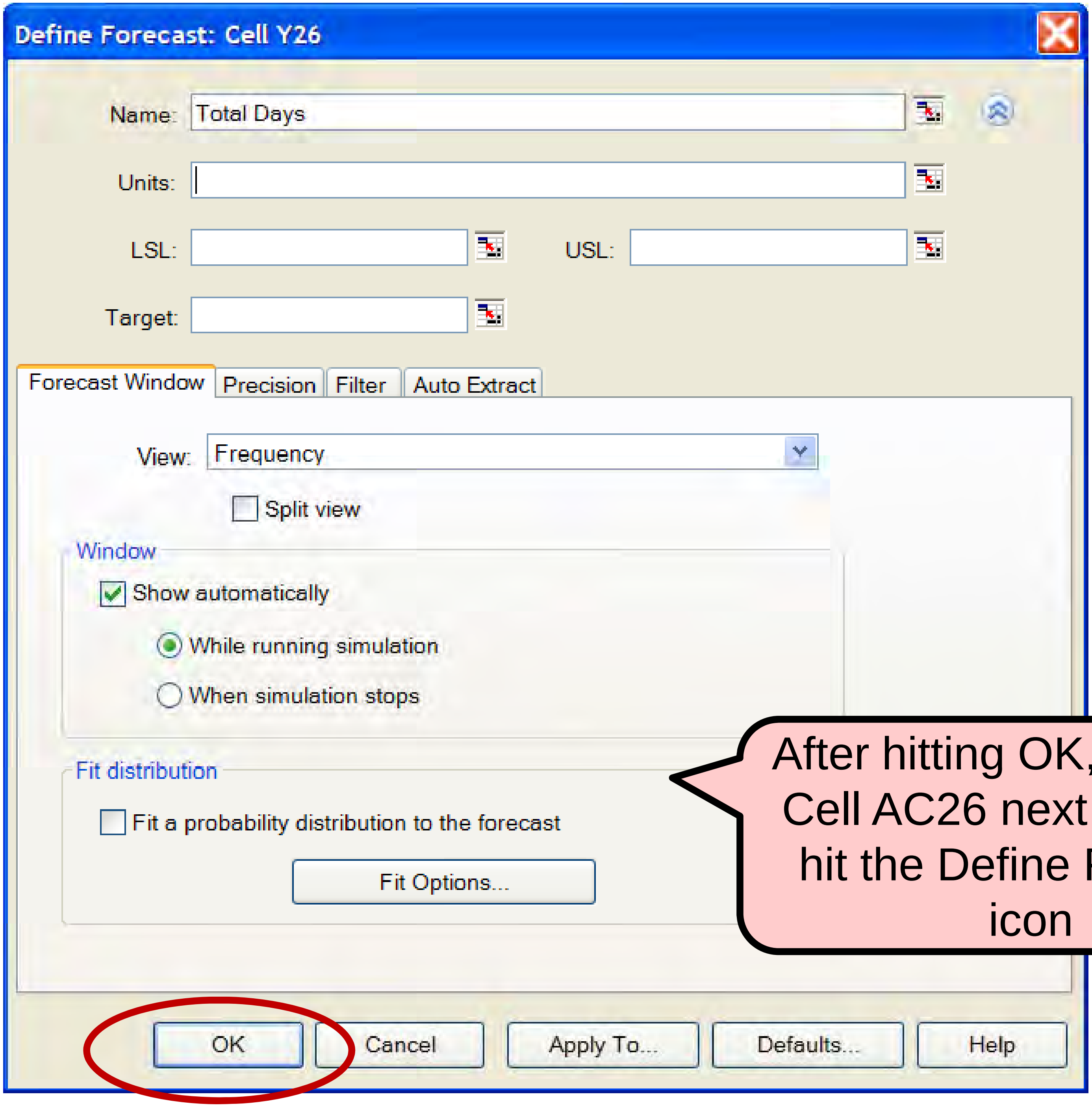

Define Assumption: Cell M24
Edit
View
Parameters
Preferences
Help
Name:
Feature 20 Budget
Triangular Distribution
Not for Commercial Use
Probability
$28.000
$30.000
$32.000
$34.000
$36.000
$38.000
$40.000
Minimum
$27.720
Likeliest
$33.264
Maximum
$39.888
OK
Cancel
Enter
Gallery
Correlate...
Help



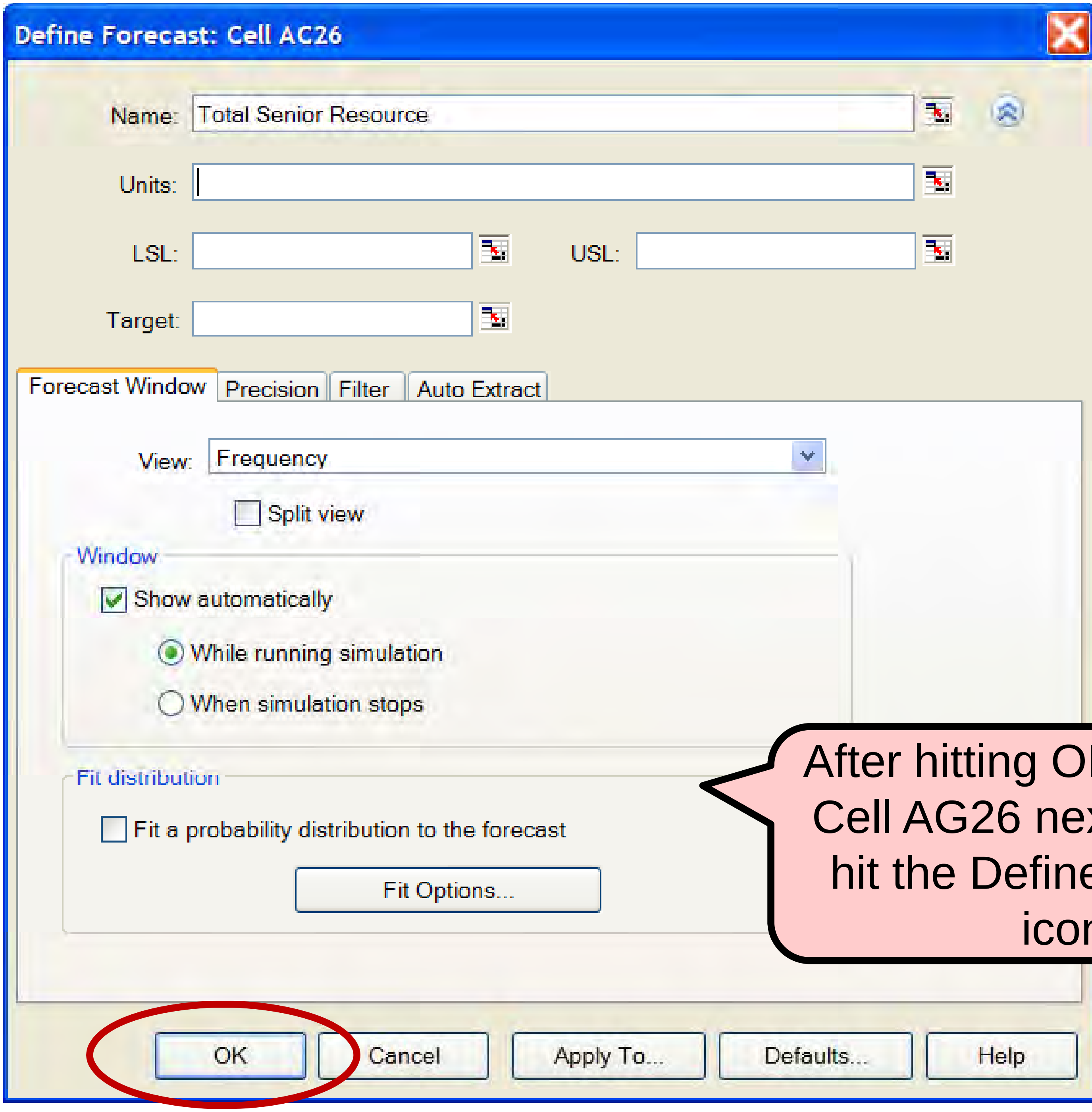

Highlight Cell O26 and then hit the Define Forecast icon



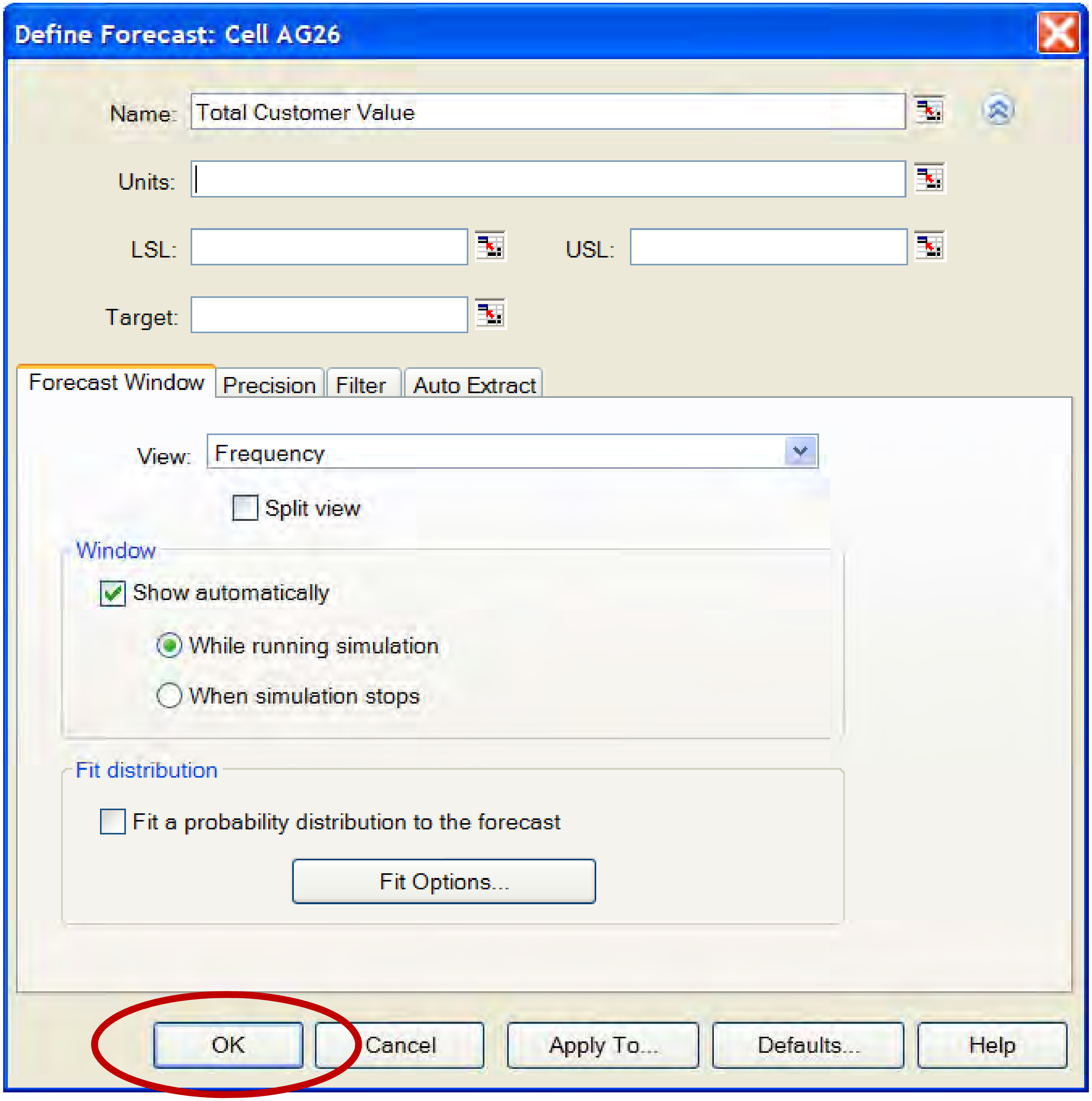

After hitting OK, Highlight Cell Y26 next and then hit the Define Forecast icon

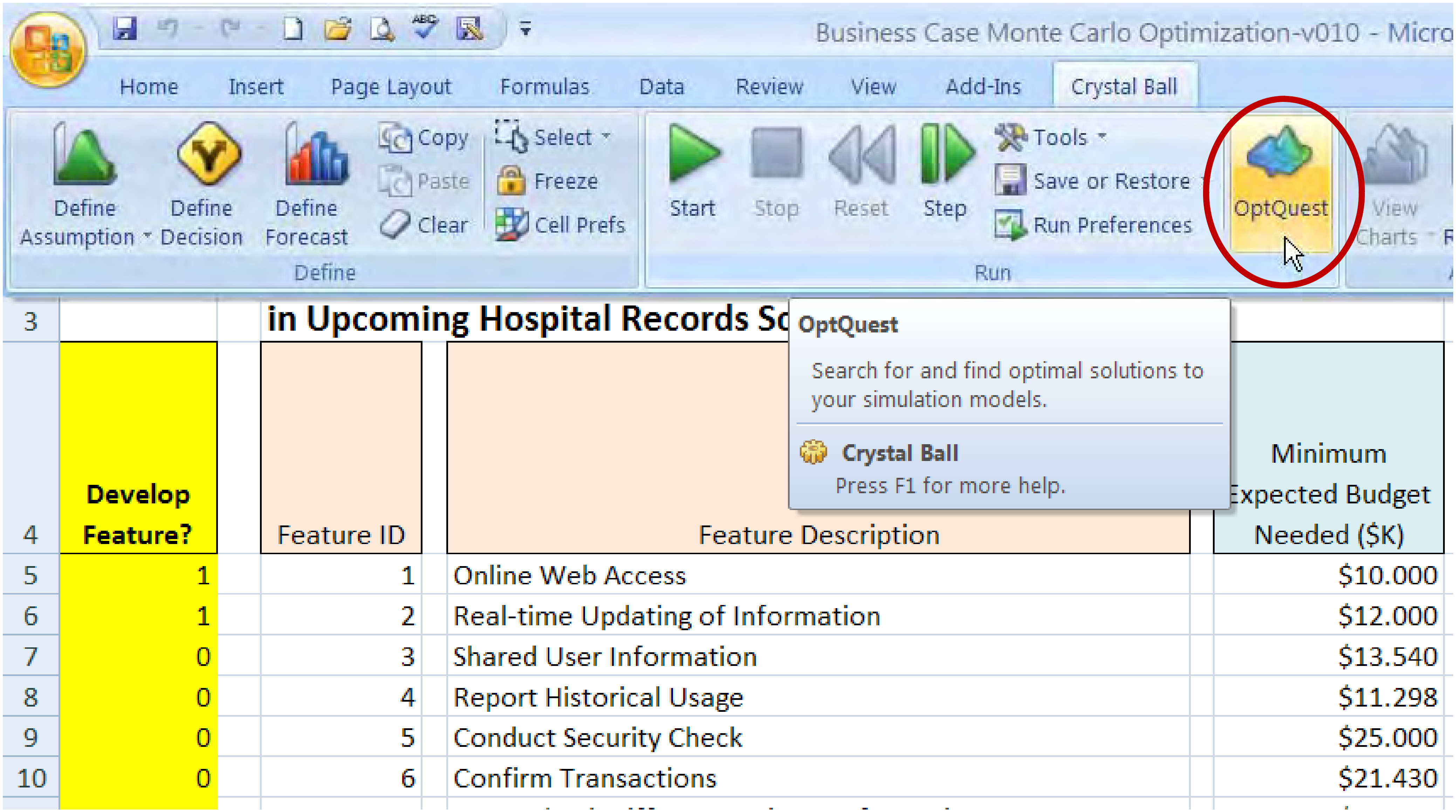


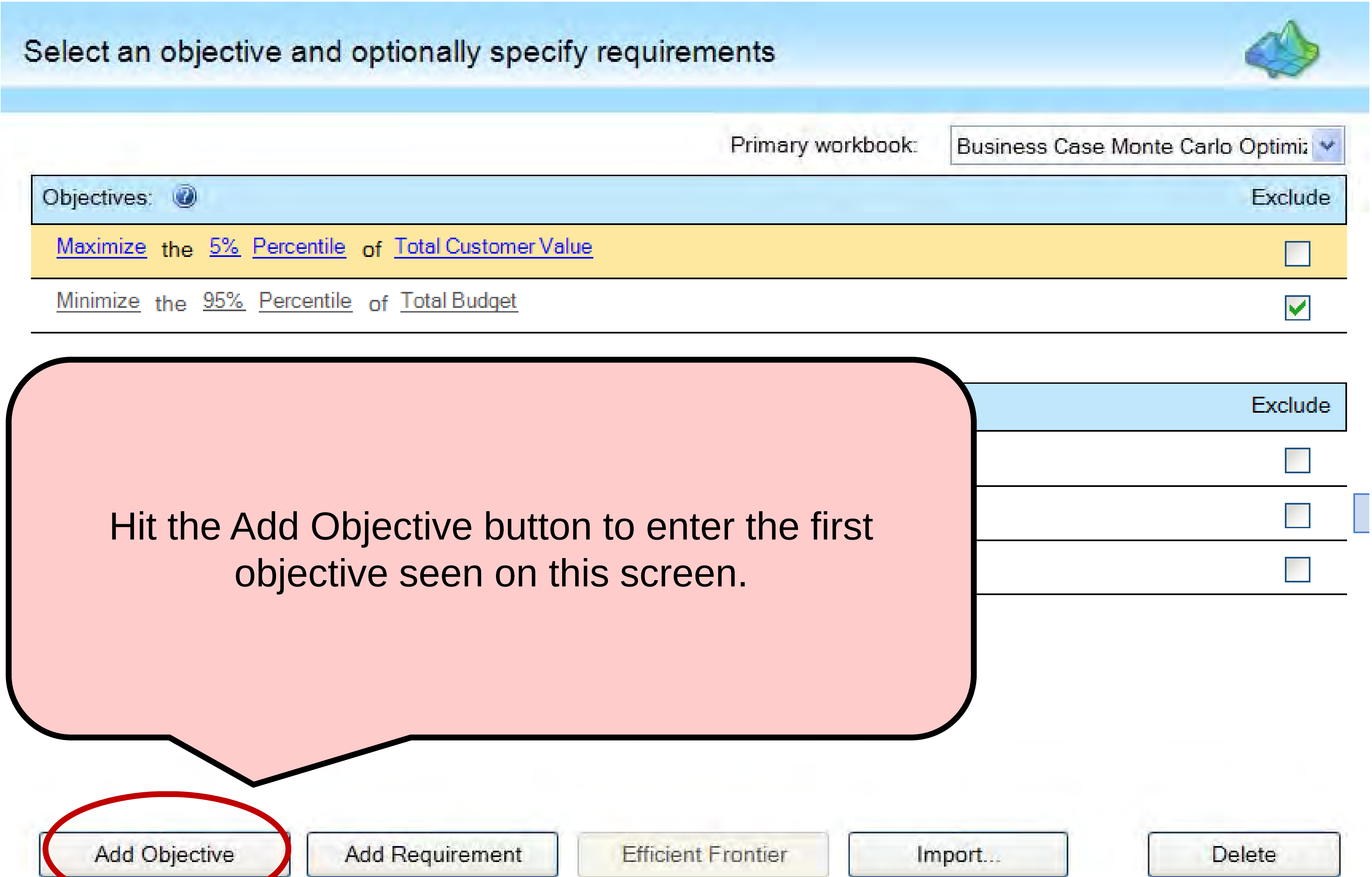

After hitting OK, Highlight Cell AC26 next and then hit the Define Forecast icon




After hitting OK, Highlight Cell AG26 next and then hit the Define Forecast icon


Define Forecast: Cell AG26
Name: Total Customer Value
Units:
LSL:
USL:
Target:
Forecast Window
Precision
Filter
Auto Extract
View: Frequency
Split view
Window
Show automatically
While running simulation
When simulation stops
Fit distribution
Fit a probability distribution to the forecast
Fit Options...
OK
Cancel
Apply To...
Defaults...
Help


Business Case Monte Carlo Optimization-v010 - Micro
Home
Insert
Page Layout
Formulas
Data
Review
View
Add-Ins
Crystal Ball
Define Assumption
Define Decision
Define Forecast
Copy
Paste
Clear
Select
Freeze
Cell Prefs
Define
Start
Stop
Reset
Step
Tools
Save or Restore
Run Preferences
Run
OptQuest
View Charts
in Upcoming Hospital Records Sc
OptQuest
Search for and find optimal solutions to your simulation models.
Crystal Ball
Press F1 for more help.
Develop Feature?
Feature ID
Feature Description
Minimum Expected Budget Needed ($K)
1 1 Online Web Access $10.000
1 2 Real-time Updating of Information $12.000
0 3 Shared User Information $13.540
0 4 Report Historical Usage $11.298
0 5 Conduct Security Check $25.000
0 6 Confirm Transactions $21.430

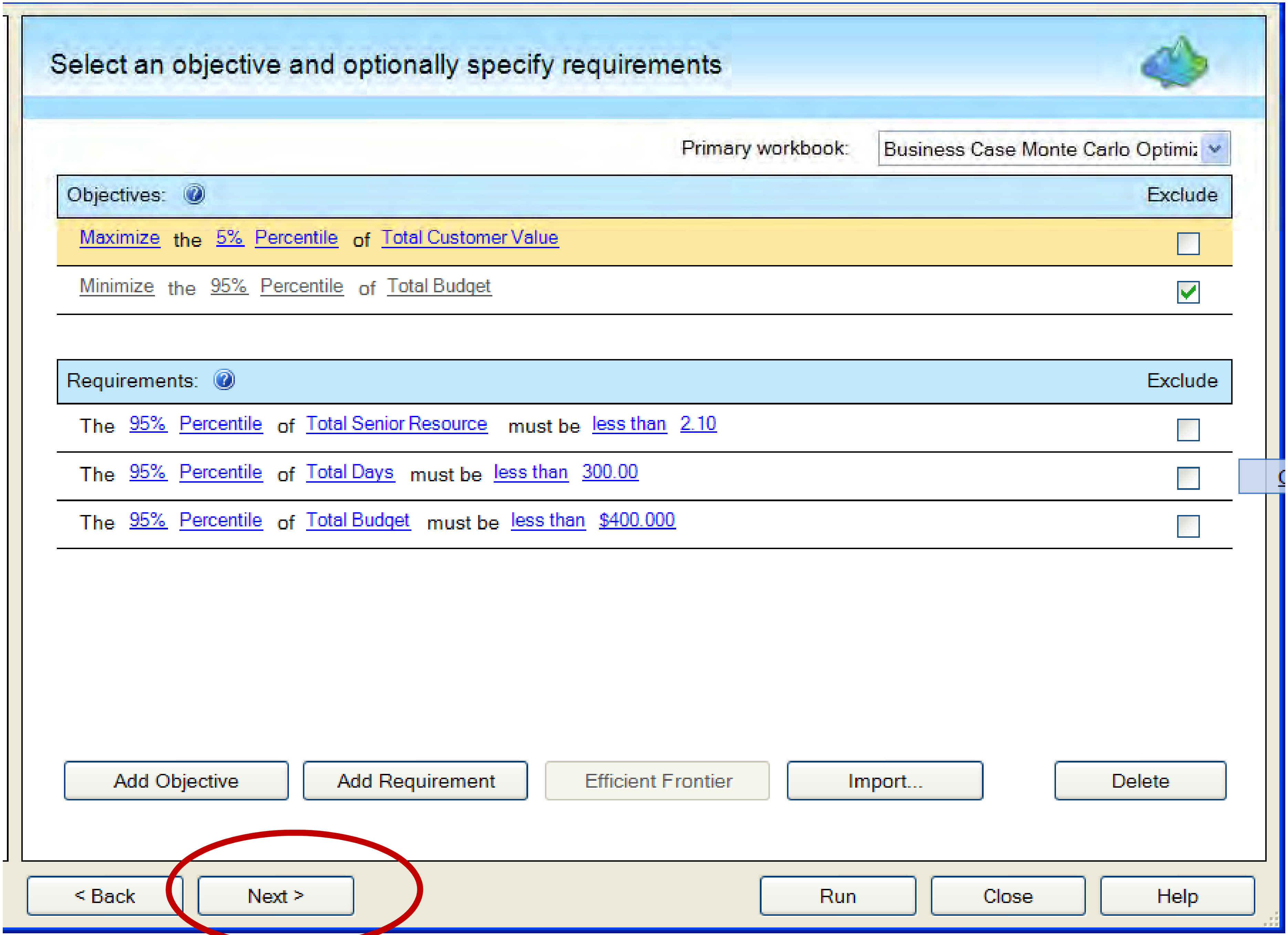

Select an objective and optionally specify requirements
Primary workbook:
Business Case Monte Carlo Optimiz
Objectives:
Exclude
Maximize the 5% Percentile of Total Customer Value
Minimize the 95% Percentile of Total Budget
Exclude
Hit the Add Objective button to enter the first objective seen on this screen.
Add Objective
Add Requirement
Efficient Frontier
Import...
Delete

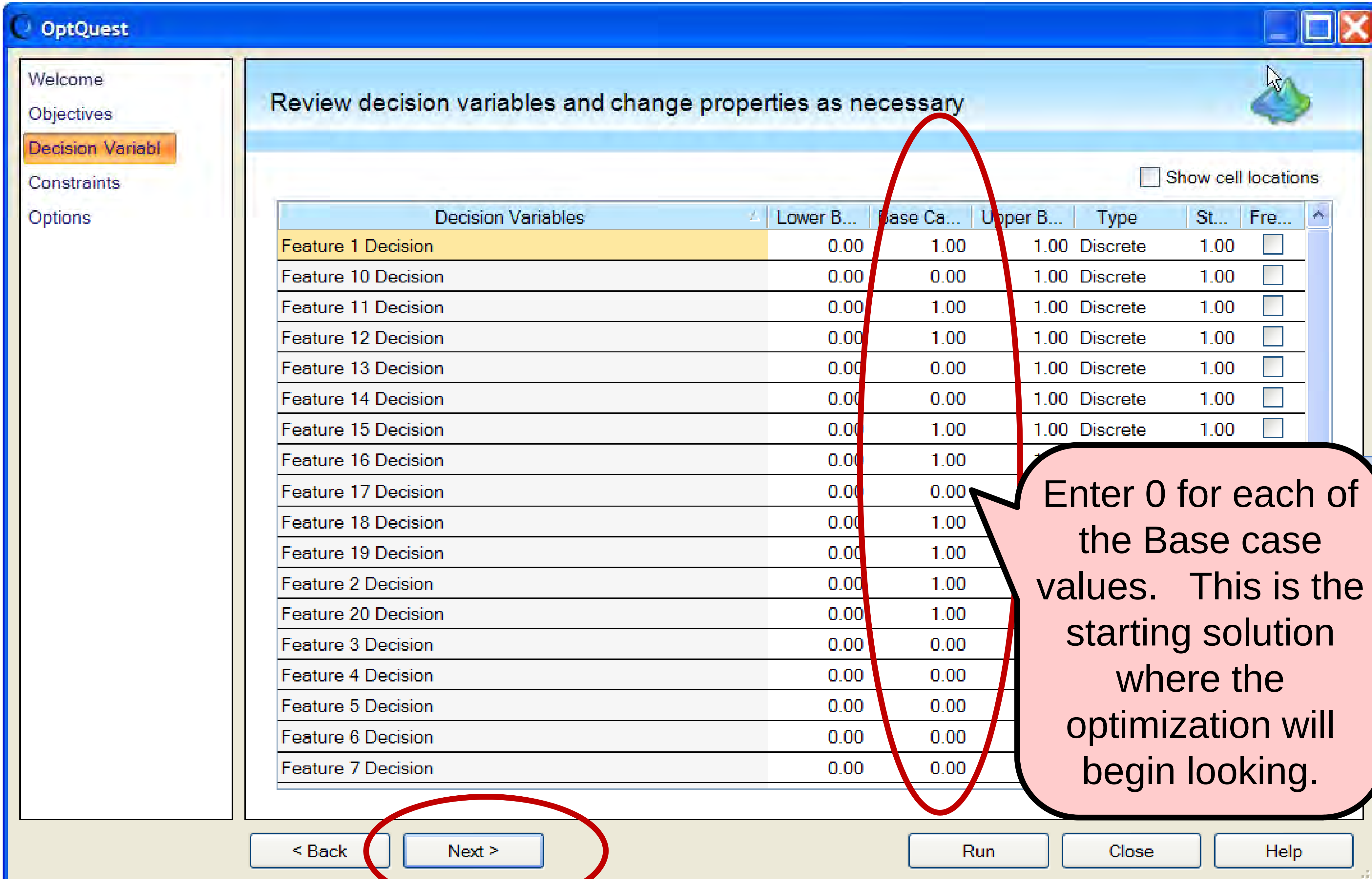
Select an objective and optionally specify requirements

Then, Hit the Add Requirement button 3 times to enter the three Requirements seen on this screen.

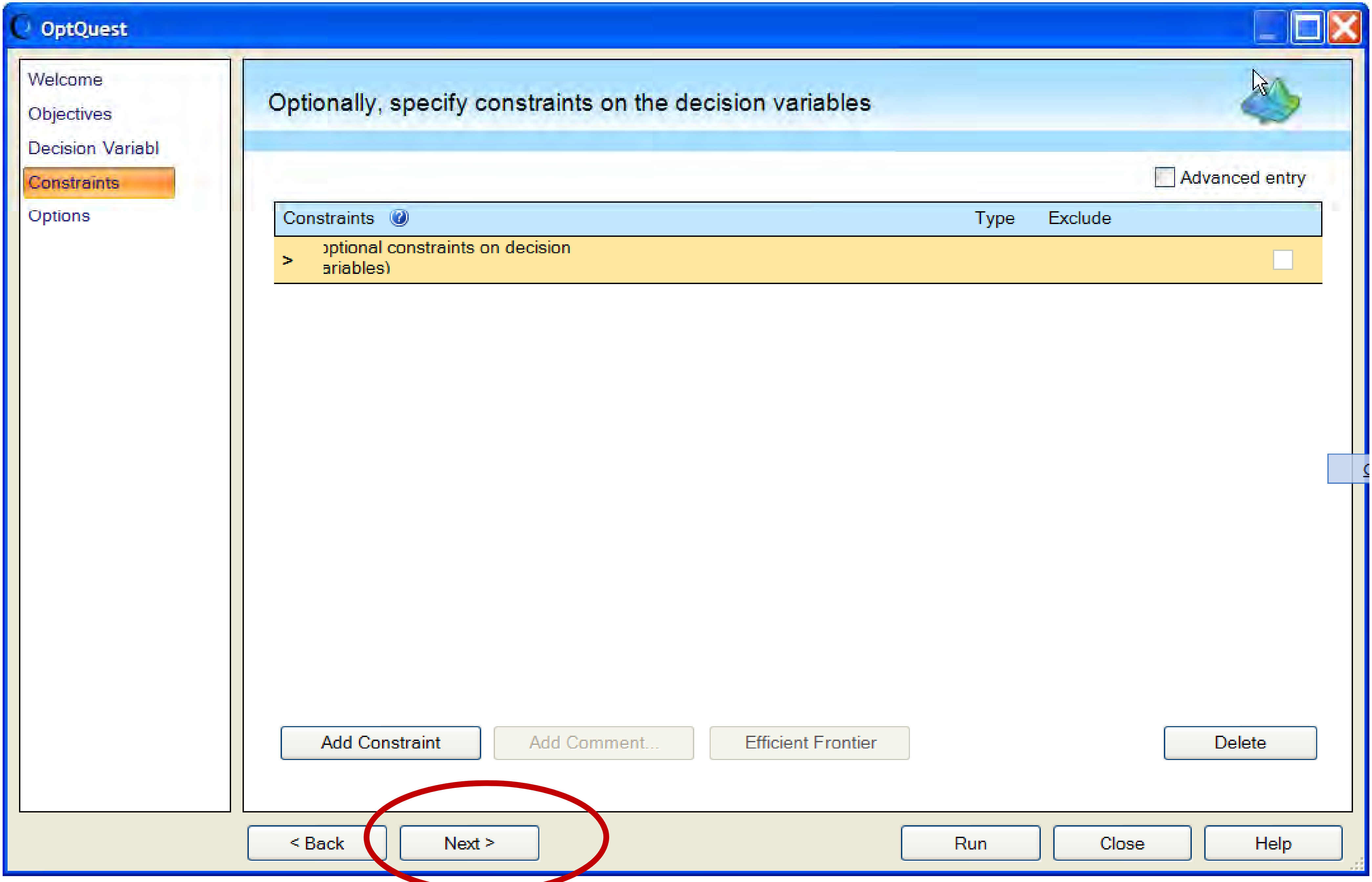
| Requirements: | Exclude |
|---|---|
| The 95% Percentile of Total Senior Resource must be less than 2.10 | ☐ |
| The 95% Percentile of Total Days must be less than 300.00 | ☐ |
| The 95% Percentile of Total Budget must be less than $400,000 | ☐ |

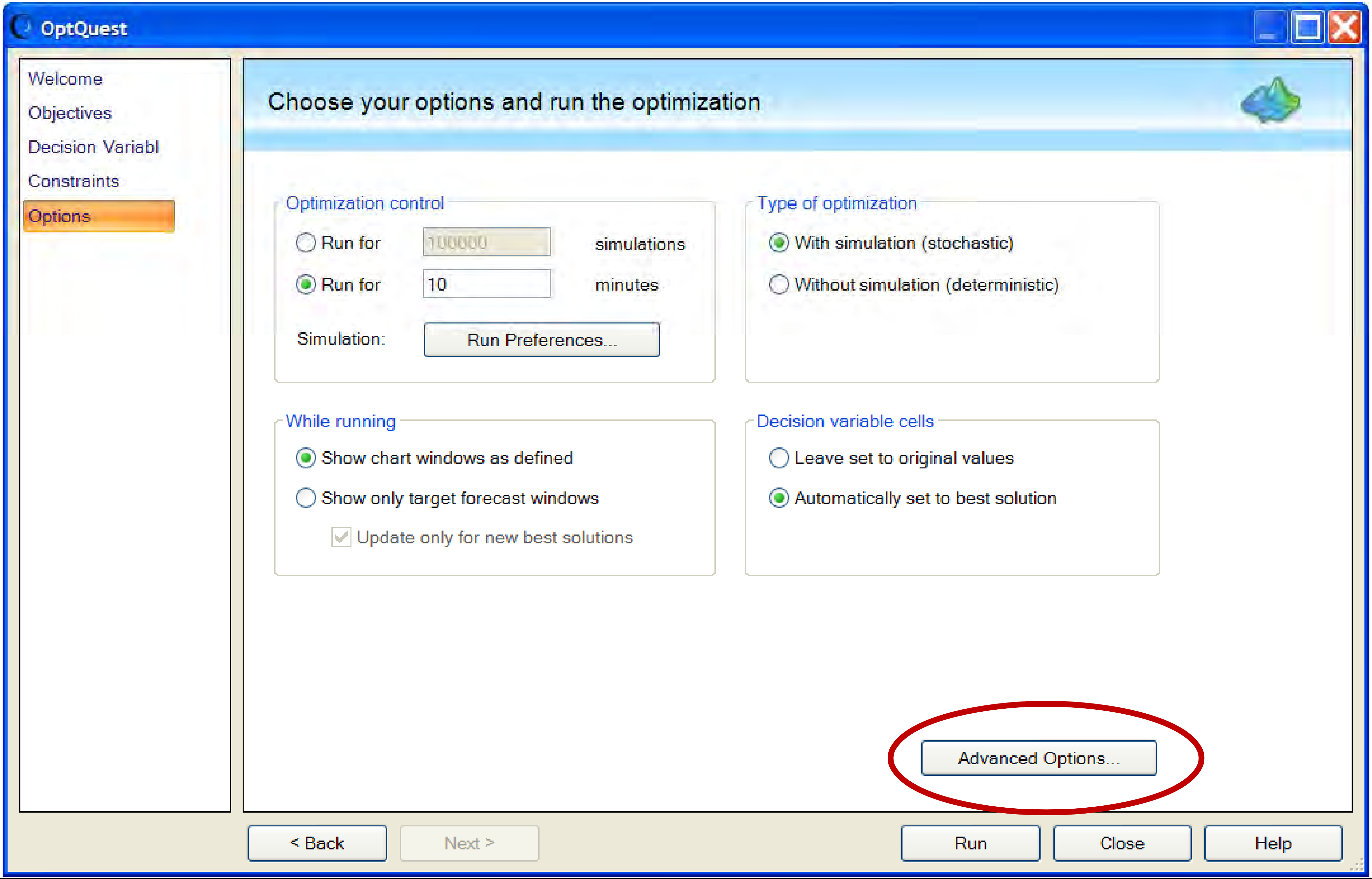
Add Objective | Add Requirement | Efficient Frontier | Import... | Delete

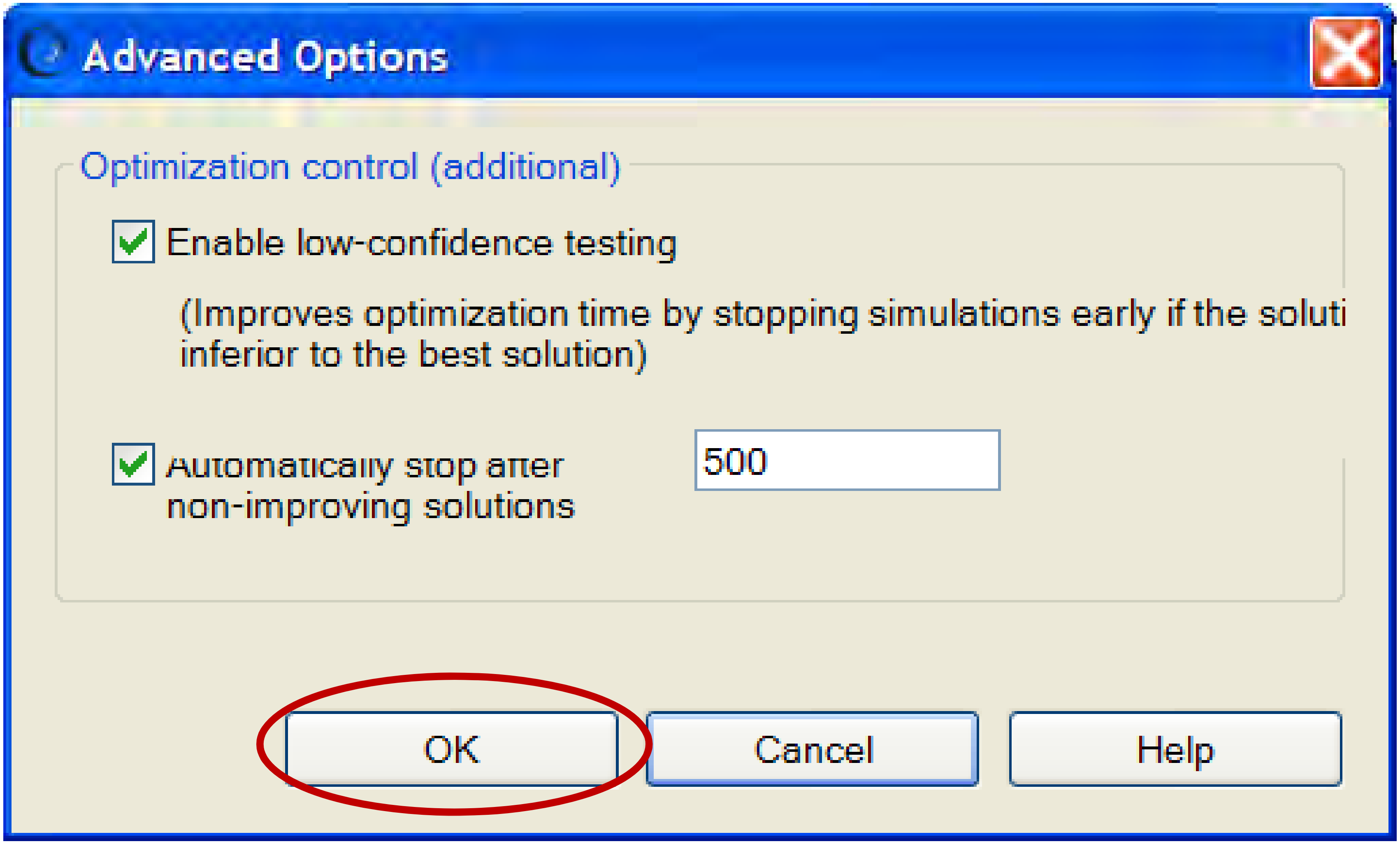

Select an objective and optionally specify requirements
Primary workbook:
Business Case Monte Carlo Optimi:
Objectives:
Exclude
Maximize the 5% Percentile of Total Customer Value
Minimize the 95% Percentile of Total Budget
Requirements:
Exclude
The 95% Percentile of Total Senior Resource must be less than 2.10
The 95% Percentile of Total Days must be less than 300.00
The 95% Percentile of Total Budget must be less than $400,000
Add Objective
Add Requirement
Efficient Frontier
Import...
Delete
< Back
Next >
Run
Close
Help

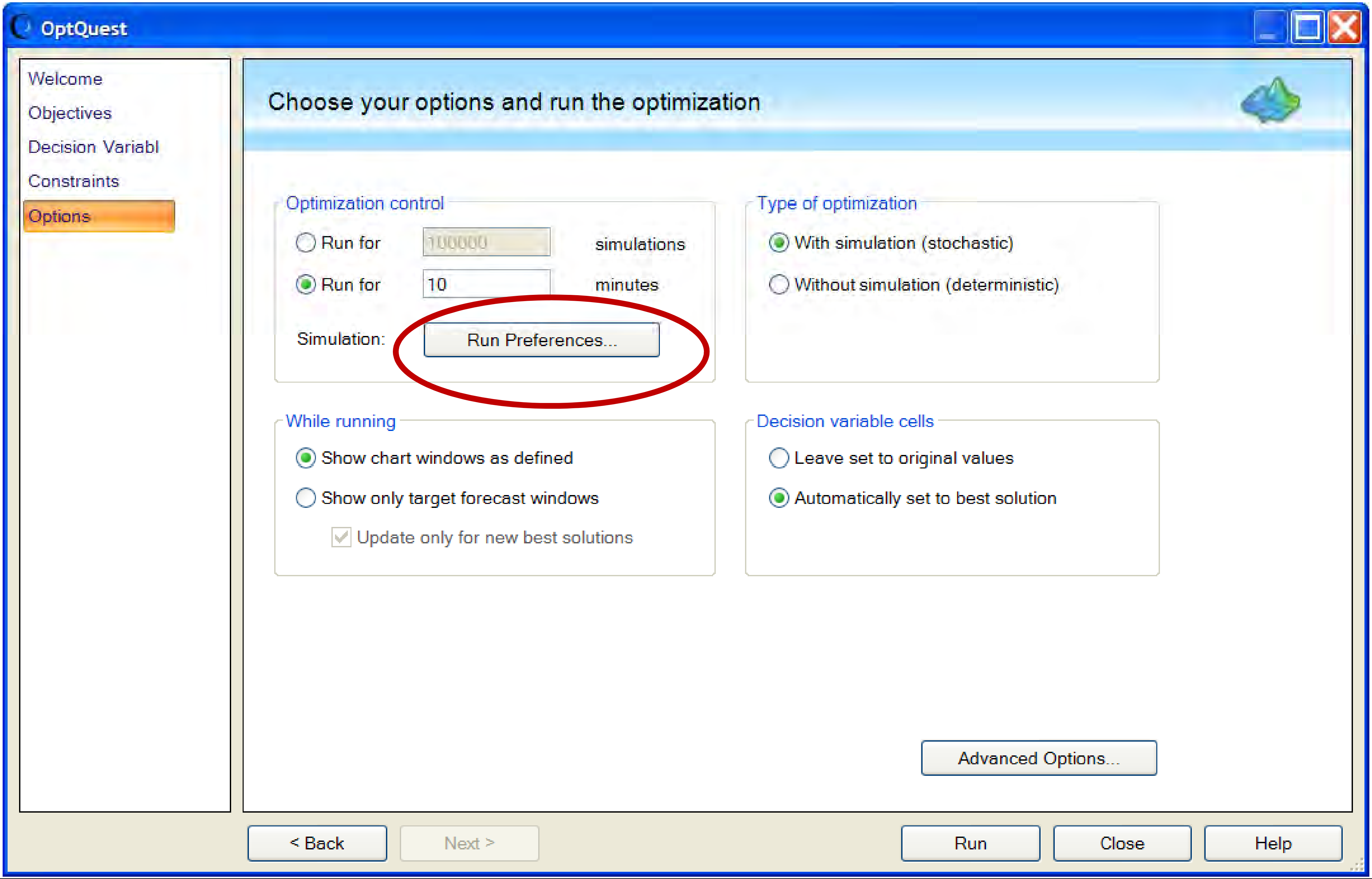


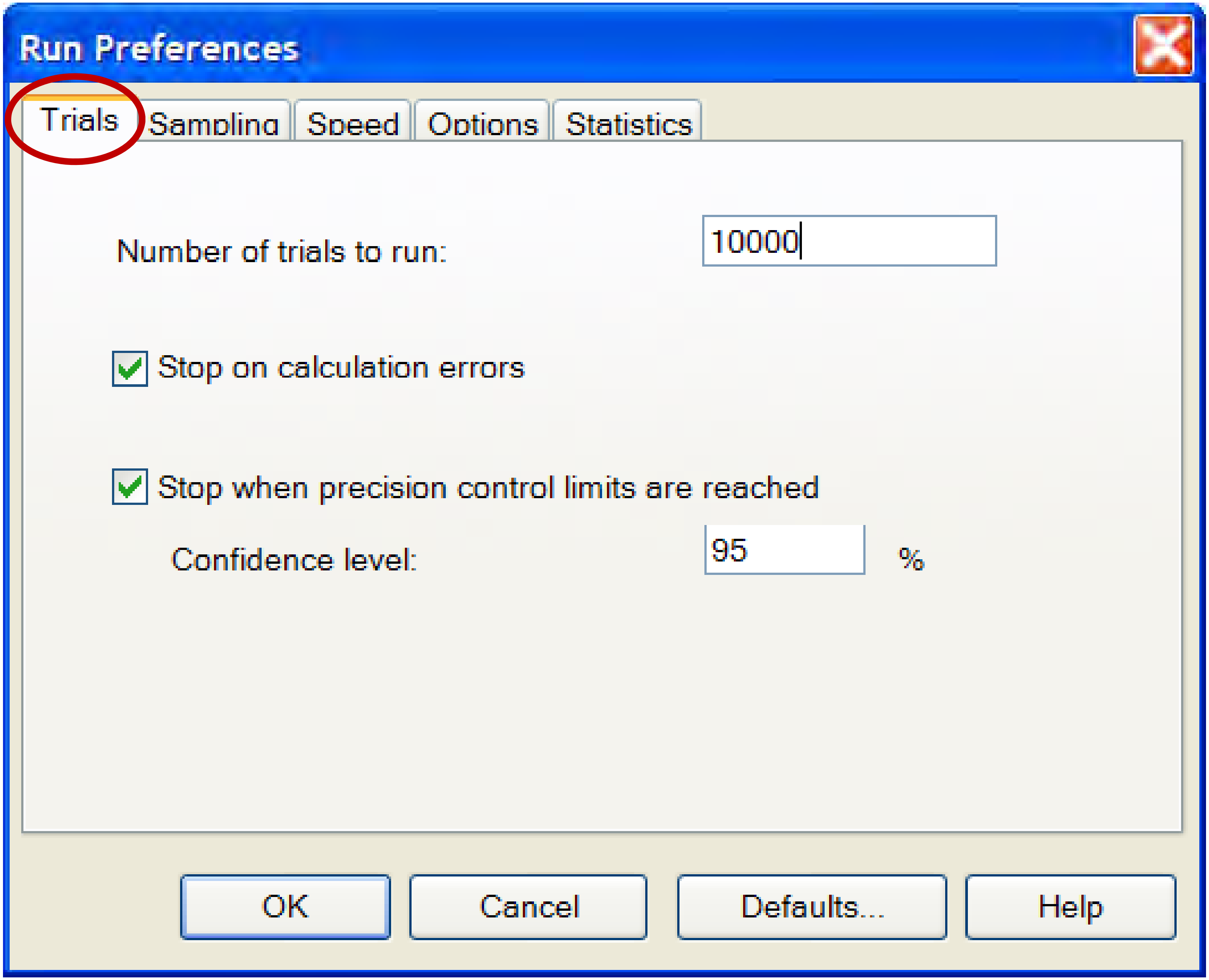

| Decision Variables | Lower B... | Base Ca... | Upper B... | Type | St... | Fre... |
|---|---|---|---|---|---|---|
| Feature 1 Decision | 0.00 | 1.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 10 Decision | 0.00 | 0.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 11 Decision | 0.00 | 1.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 12 Decision | 0.00 | 1.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 13 Decision | 0.00 | 0.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 14 Decision | 0.00 | 0.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 15 Decision | 0.00 | 1.00 | 1.00 | Discrete | 1.00 | ☐ |
| Feature 16 Decision | 0.00 | 1.00 | | | | |
| Feature 17 Decision | 0.00 | 0.00 | | | | |
| Feature 18 Decision | 0.00 | 1.00 | | | | |
| Feature 19 Decision | 0.00 | 1.00 | | | | |
| Feature 2 Decision | 0.00 | 1.00 | | | | |
| Feature 20 Decision | 0.00 | 1.00 | | | | |
| Feature 3 Decision | 0.00 | 0.00 | | | | |
| Feature 4 Decision | 0.00 | 0.00 | | | | |
| Feature 5 Decision | 0.00 | 0.00 | | | | |
| Feature 6 Decision | 0.00 | 0.00 | | | | |
| Feature 7 Decision | 0.00 | 0.00 | | | | |

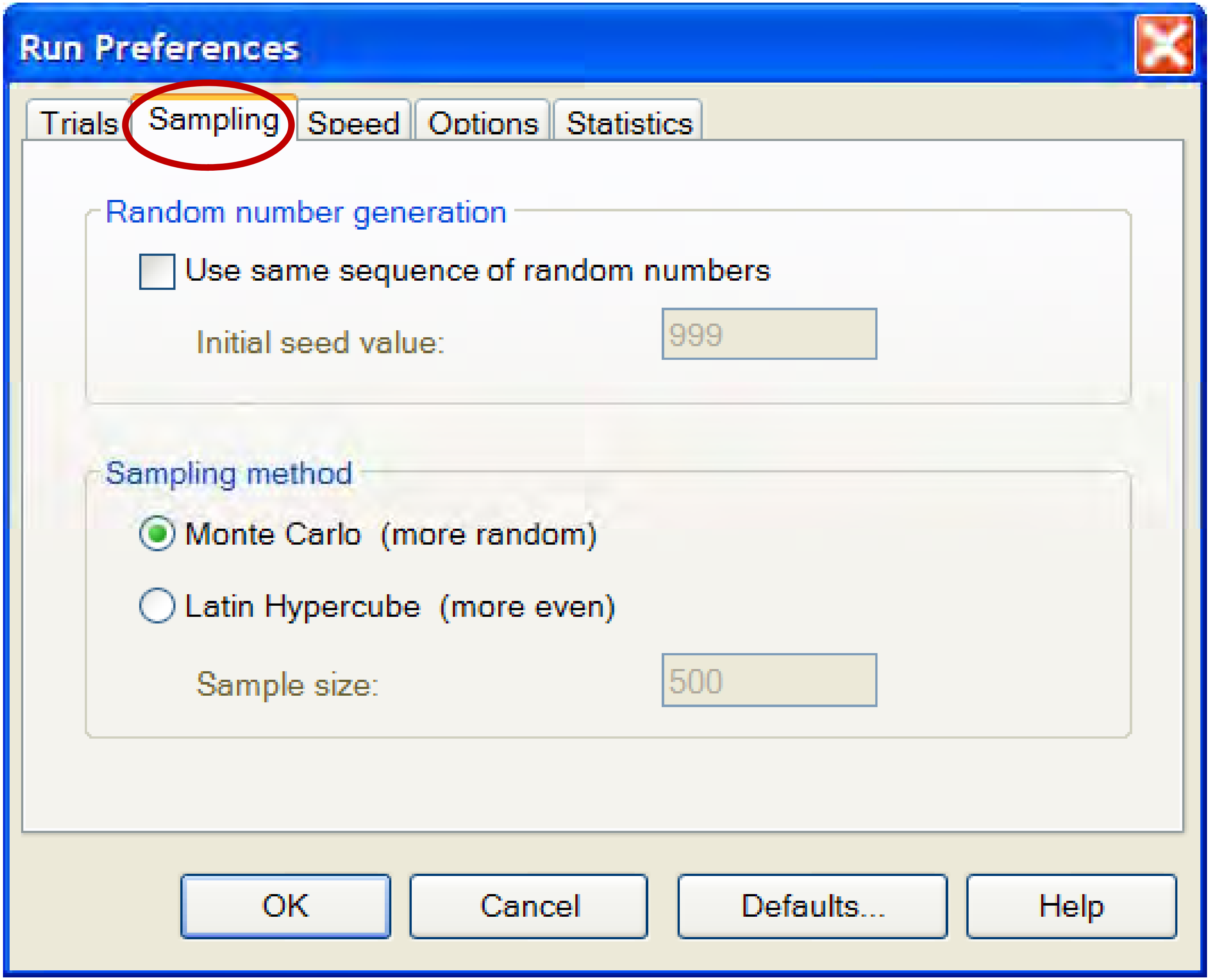

OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Optionally, specify constraints on the decision variables
Advanced entry
Constraints
Type
Exclude
>
ɔptional constraints on decision
ariables)
Add Constraint
Add Comment...
Efficient Frontier
Delete
< Back
Next >
Run
Close
Help

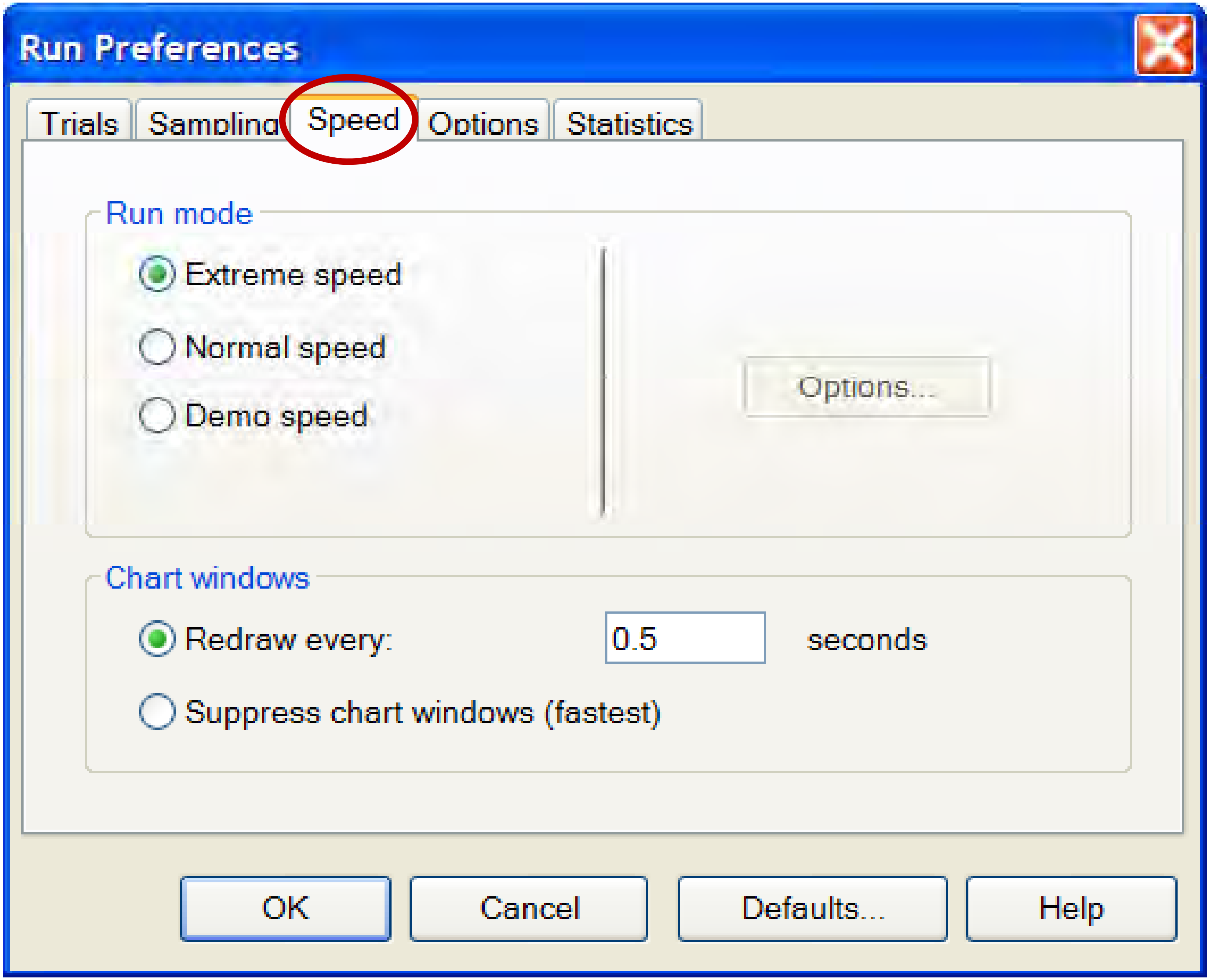

OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Choose your options and run the optimization
Optimization control
Run for
100000
simulations
Run for
10
minutes
Simulation:
Run Preferences...
Type of optimization
With simulation (stochastic)
Without simulation (deterministic)
While running
Show chart windows as defined
Show only target forecast windows
Update only for new best solutions
Decision variable cells
Leave set to original values
Automatically set to best solution
Advanced Options...
< Back
Next >
Run
Close
Help

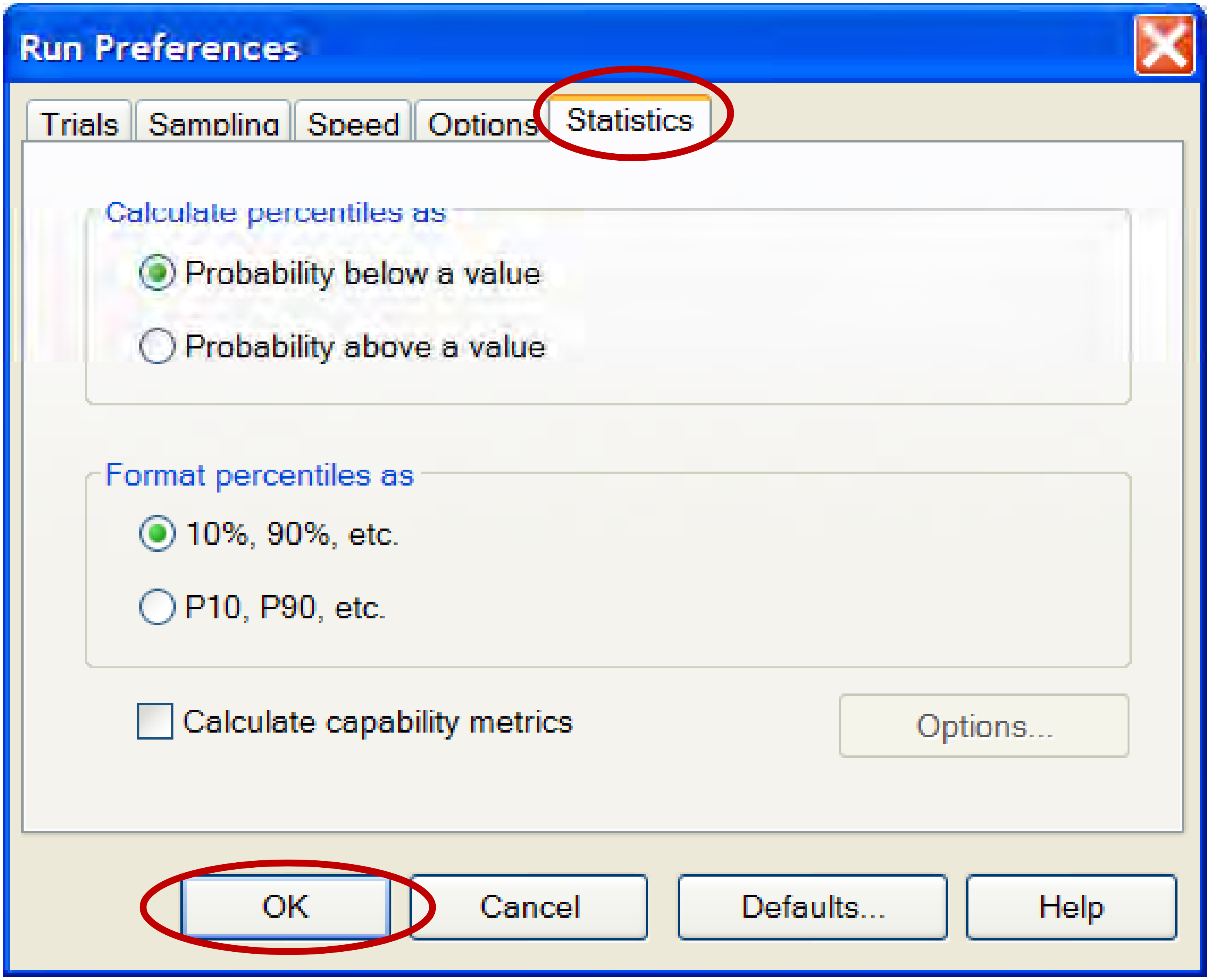

Advanced Options
Optimization control (additional)
Enable low-confidence testing
(Improves optimization time by stopping simulations early if the soluti
inferior to the best solution)
Automatically stop after
non-improving solutions
500
OK
Cancel
Help

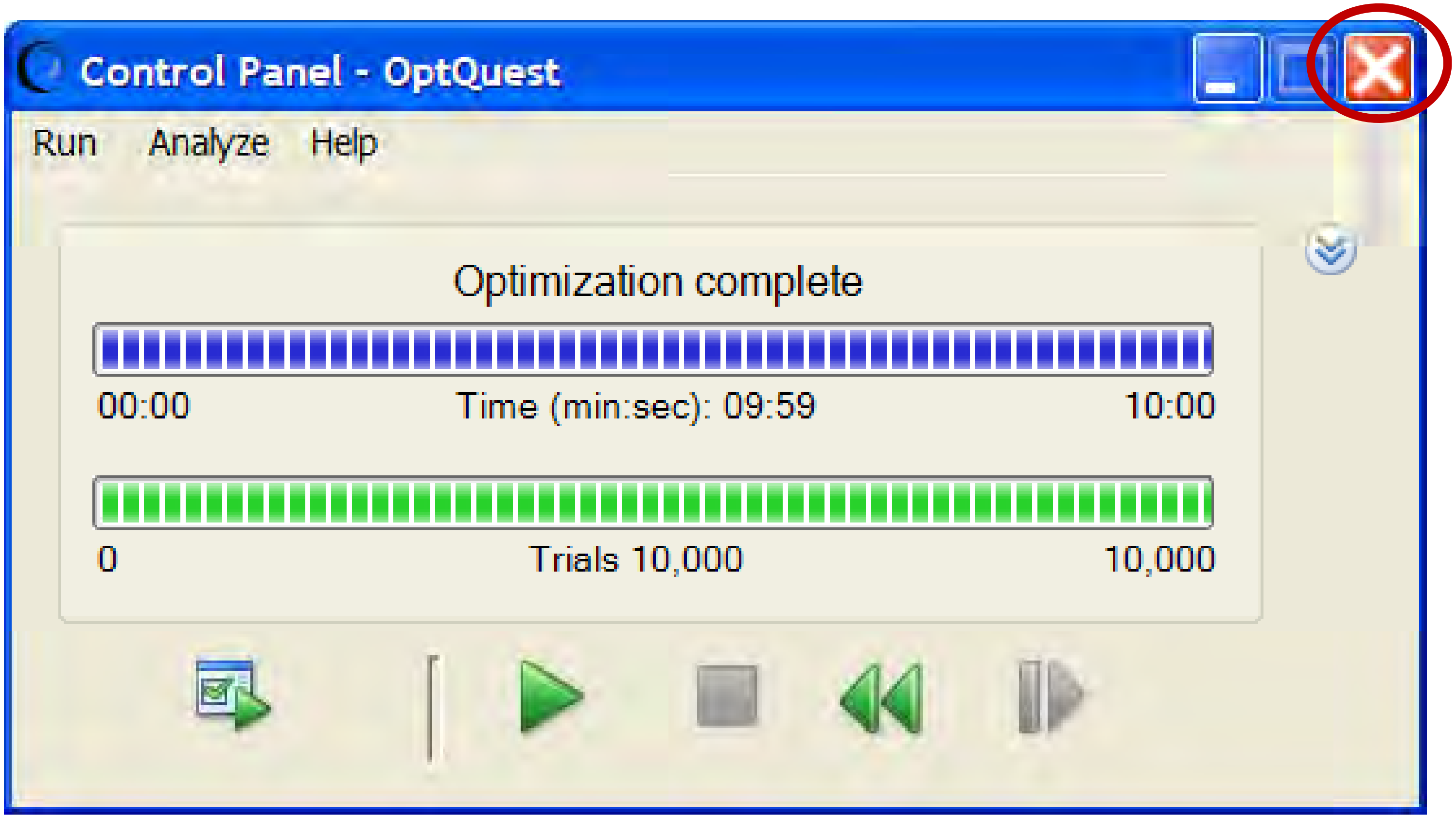

OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Choose your options and run the optimization
Optimization control
Run for
100000
simulations
Run for
10
minutes
Simulation:
Run Preferences...
Type of optimization
With simulation (stochastic)
Without simulation (deterministic)
While running
Show chart windows as defined
Show only target forecast windows
Update only for new best solutions
Decision variable cells
Leave set to original values
Automatically set to best solution
Advanced Options...
< Back
Next >
Run
Close
Help

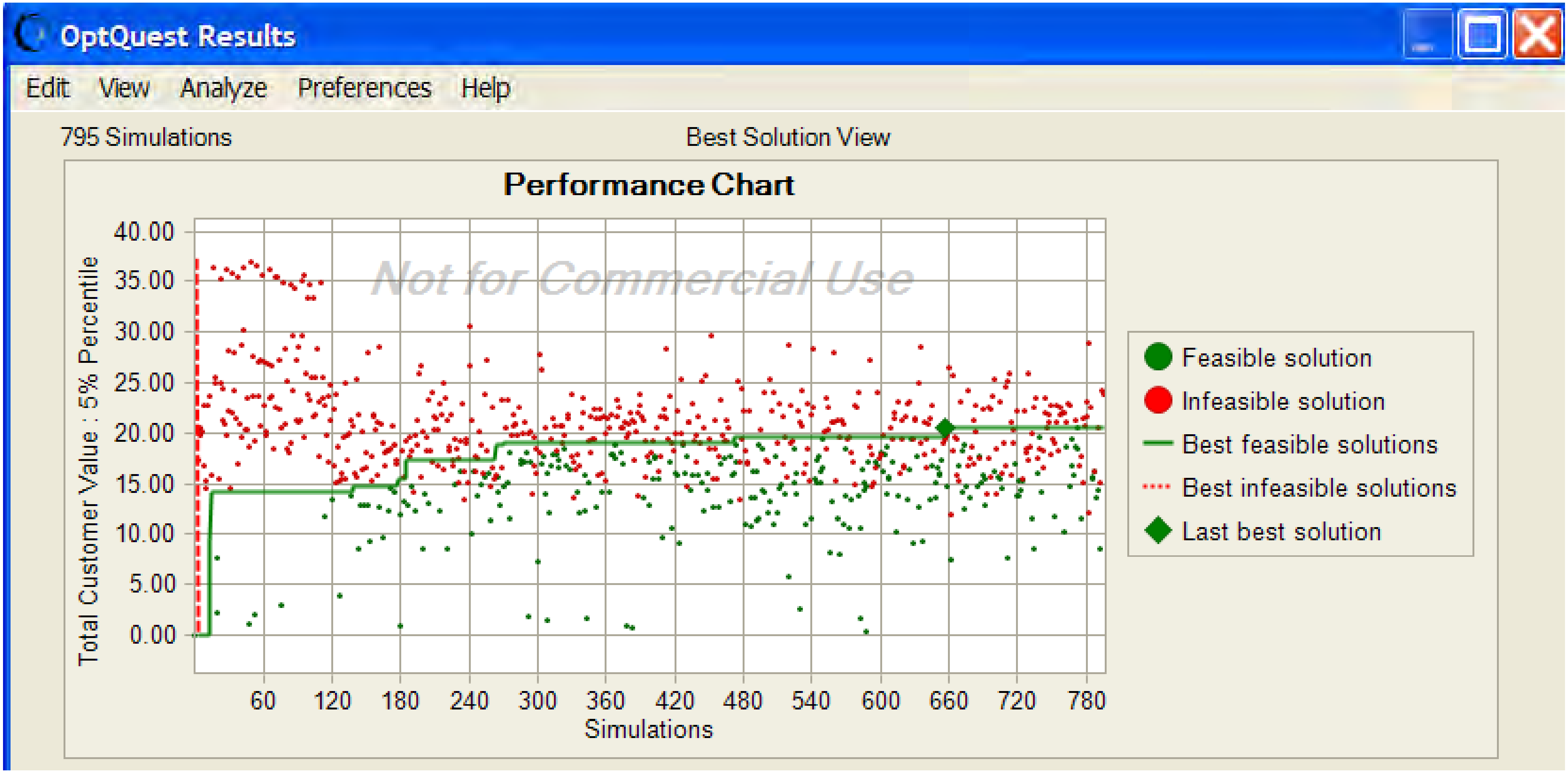

Run Preferences
Trials
Sampling
Speed
Options
Statistics
Number of trials to run:
10000
Stop on calculation errors
Stop when precision control limits are reached
Confidence level:
95
%
OK
Cancel
Defaults...
Help


Run Preferences
Trials
Sampling
Speed
Options
Statistics
Random number generation
Use same sequence of random numbers
Initial seed value:
999
Sampling method
Monte Carlo (more random)
Latin Hypercube (more even)
Sample size:
500
OK
Cancel
Defaults...
Help


Run Preferences
Trials
Sampling
Speed
Options
Statistics
Run mode
Extreme speed
Normal speed
Demo speed
Options...
Chart windows
Redraw every:
0.5
seconds
Suppress chart windows (fastest)
OK
Cancel
Defaults...
Help


Run Preferences
Trials
Sampling
Speed
Options
Statistics
Calculate percentiles as
Probability below a value
Probability above a value
Format percentiles as
10%, 90%, etc.
P10, P90, etc.
Calculate capability metrics
Options...
OK
Cancel
Defaults...
Help


Control Panel - OptQuest
Run
Analyze
Help
Optimization complete
00:00
Time (min:sec): 09:59
10:00
0
Trials 10,000
10,000

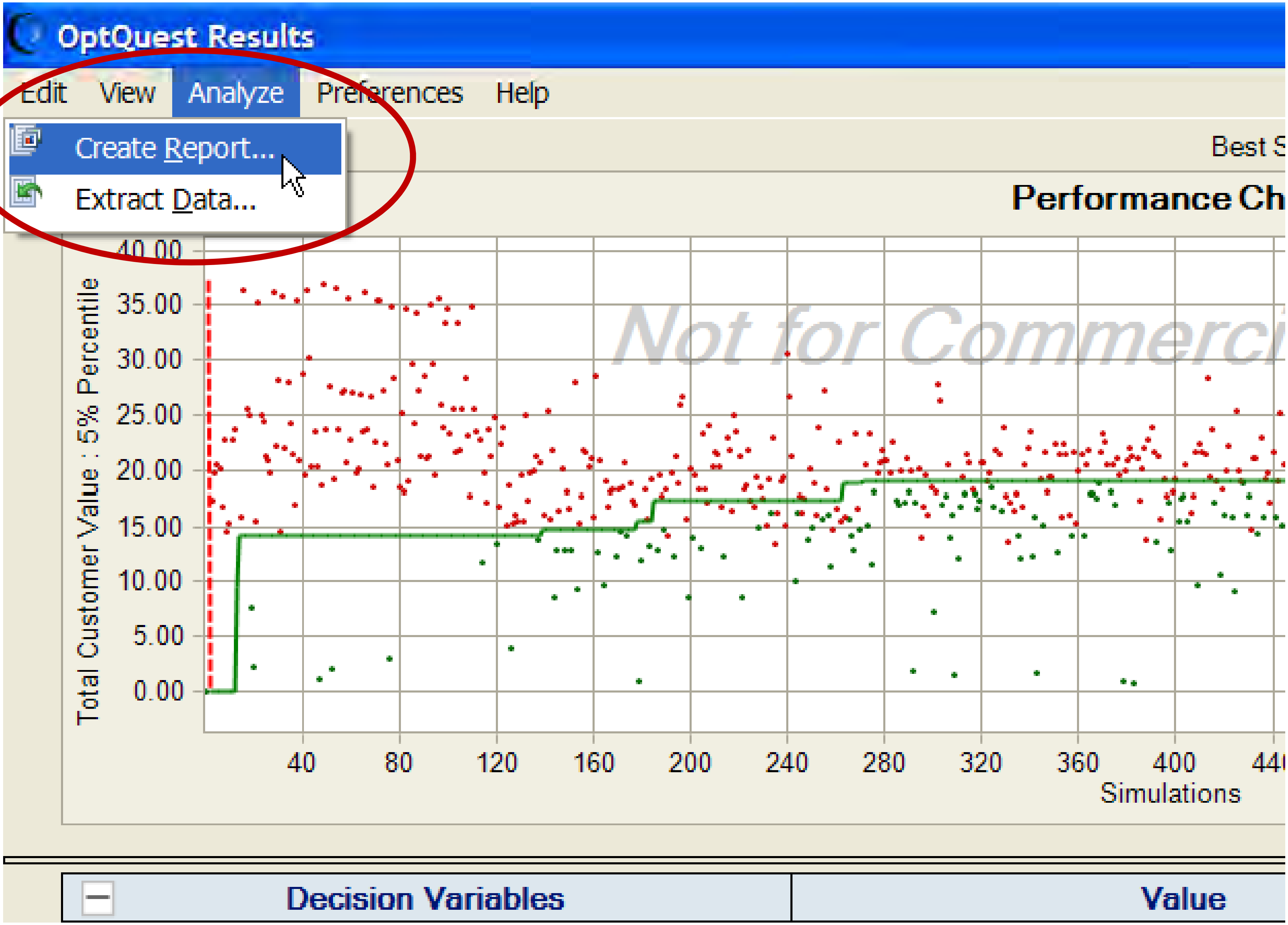

OptQuest Results
Edit View Analyze Preferences Help
795 Simulations
Best Solution View
Performance Chart
Not for Commercial Use
Total Customer Value : 5% Percentile
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
60
120
180
240
300
360
420
480
540
600
660
720
780
Simulations
Feasible solution
Infeasible solution
Best feasible solutions
Best infeasible solutions
Last best solution

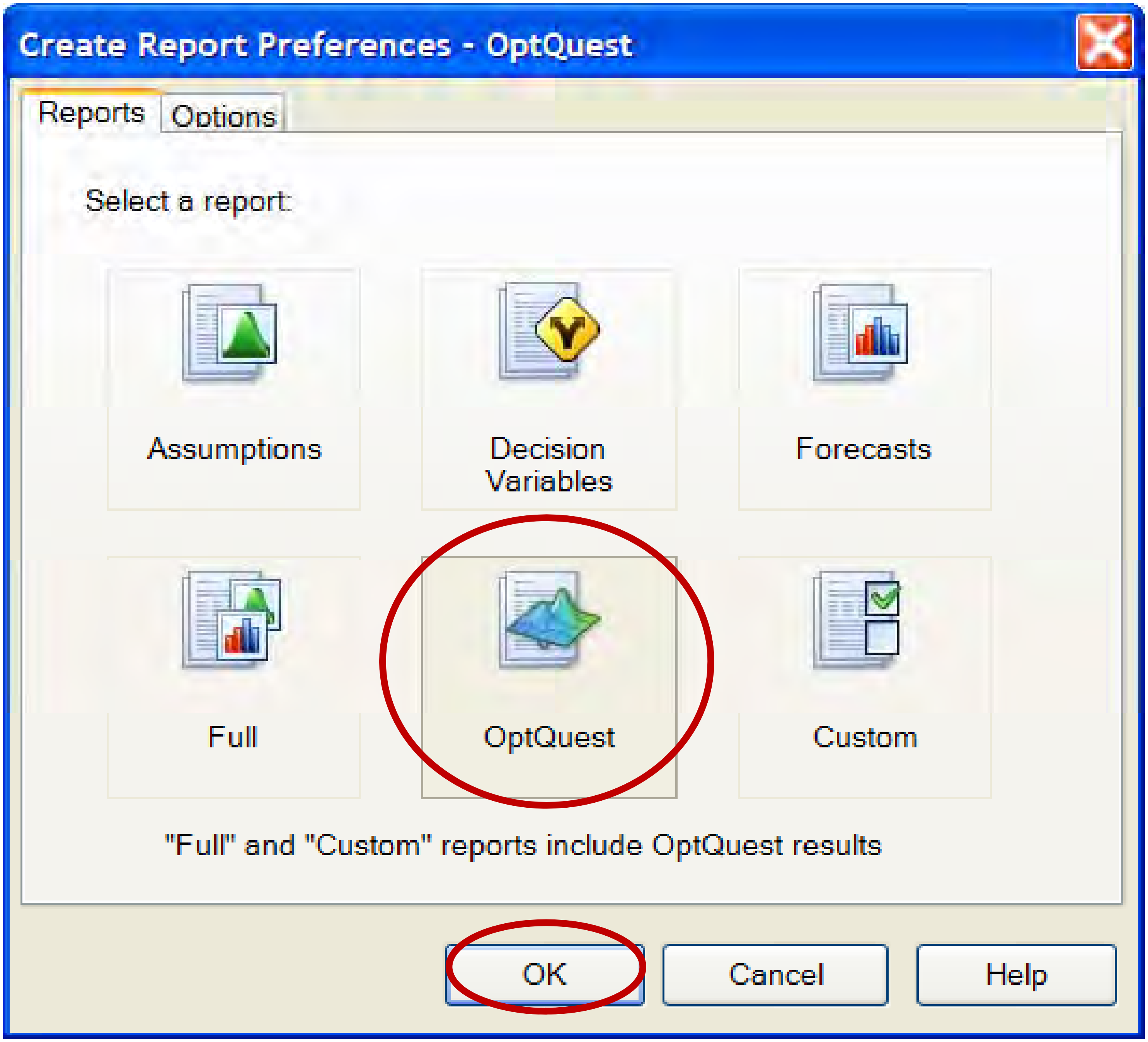
Best Solution: Simulation # 656

| Objectives | Value |
| --- | --- |
| Maximize the 5% Percentile of Total Customer Value | 20.54 |

| Requirements | Value |
| --- | --- |
| The 95% Percentile of Total Senior Resource must b... | 1.57 |
| The 95% Percentile of Total Days must be less than ... | 296.10 |
| The 95% Percentile of Total Budget must be less tha... | $359.952 |

| Decision Variables | Value |
|---|---|
| Feature 1 Decision | 0.00 |
| Feature 10 Decision | 1.00 |
| Feature 11 Decision | 0.00 |
| Feature 12 Decision | 1.00 |
| Feature 13 Decision | 1.00 |
| Feature 14 Decision | 0.00 |
| Feature 15 Decision | 1.00 |
| Feature 16 Decision | 1.00 |
| Feature 17 Decision | 1.00 |
| Feature 18 Decision | 0.00 |
| Feature 19 Decision | 1.00 |
| Feature 2 Decision | 1.00 |
| Feature 20 Decision | 1.00 |
| Feature 3 Decision | 0.00 |
| Feature 4 Decision | 0.00 |
| Feature 5 Decision | 0.00 |
| Feature 6 Decision | 0.00 |
| Feature 7 Decision | 0.00 |
| Feature 8 Decision | 0.00 |
| Feature 9 Decision | 0.00 |


OptQuest Results
Edit
View
Analyze
Preferences
Help
Create Report...
Extract Data...
Best S
Performance Ch
Not for Commerci
Total Customer Value : 5% Percentile
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
40
80
120
160
200
240
280
320
360
400
44
Simulations
Decision Variables
Value


Create Report Preferences - OptQuest
Reports
Options
Select a report:
Assumptions
Decision Variables
Forecasts
Full
OptQuest
Custom
"Full" and "Custom" reports include OptQuest results
OK
Cancel
Help

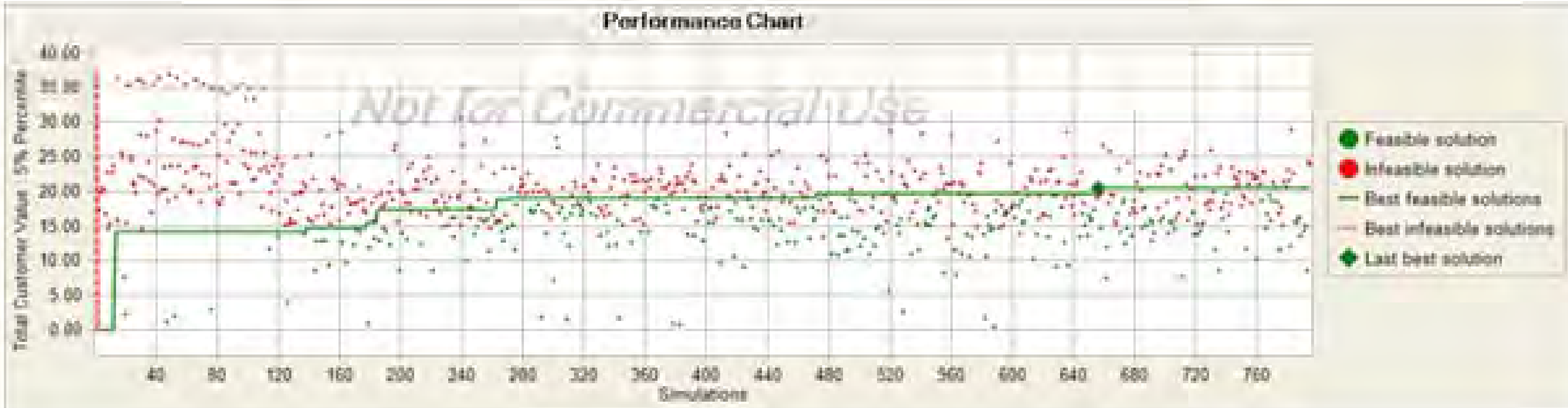
# Crystal Ball Report - OptQuest

Optimization started on 8/21/2009 at 9:20:35
Optimization stopped on 8/21/2009 at 9:30:36

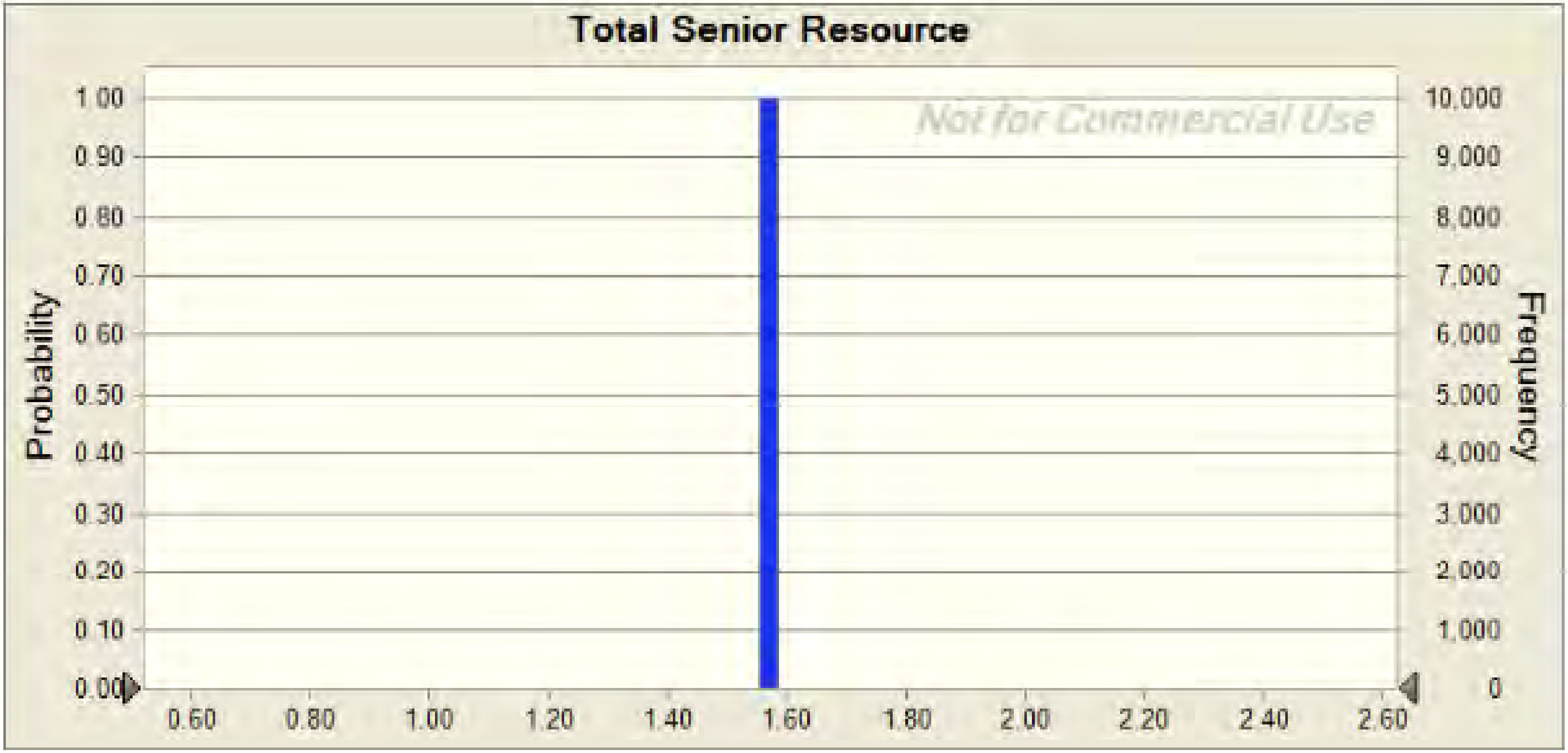
| | |
|---|---|
| **Run preferences:** | |
| Stochastic optimization (with simulation) | |
| Low-confidence testing on | |
| Maximum trials per simulation | 10,000 |
| Monte Carlo | |
| Random seed | |
| Precision control on | |
| Confidence level | 95.00% |
| **Run statistics:** | |
| Total optimization time (min:sec) | 10:01 |
| Number of simulations | 795 |
| Stopped by | |
| Trials limit reached | 406 |
| Precision control | 0 |
| Low-confidence testing | 389 |
| Infeasible constraints | 0 |
| Simulation/second (average) | 1 |
| **Other statistics:** | |
| Number of infeasible solutions | 537 |
| Due to requirements | 537 |
| Due to non-linear constraints | 0 |

Summary:

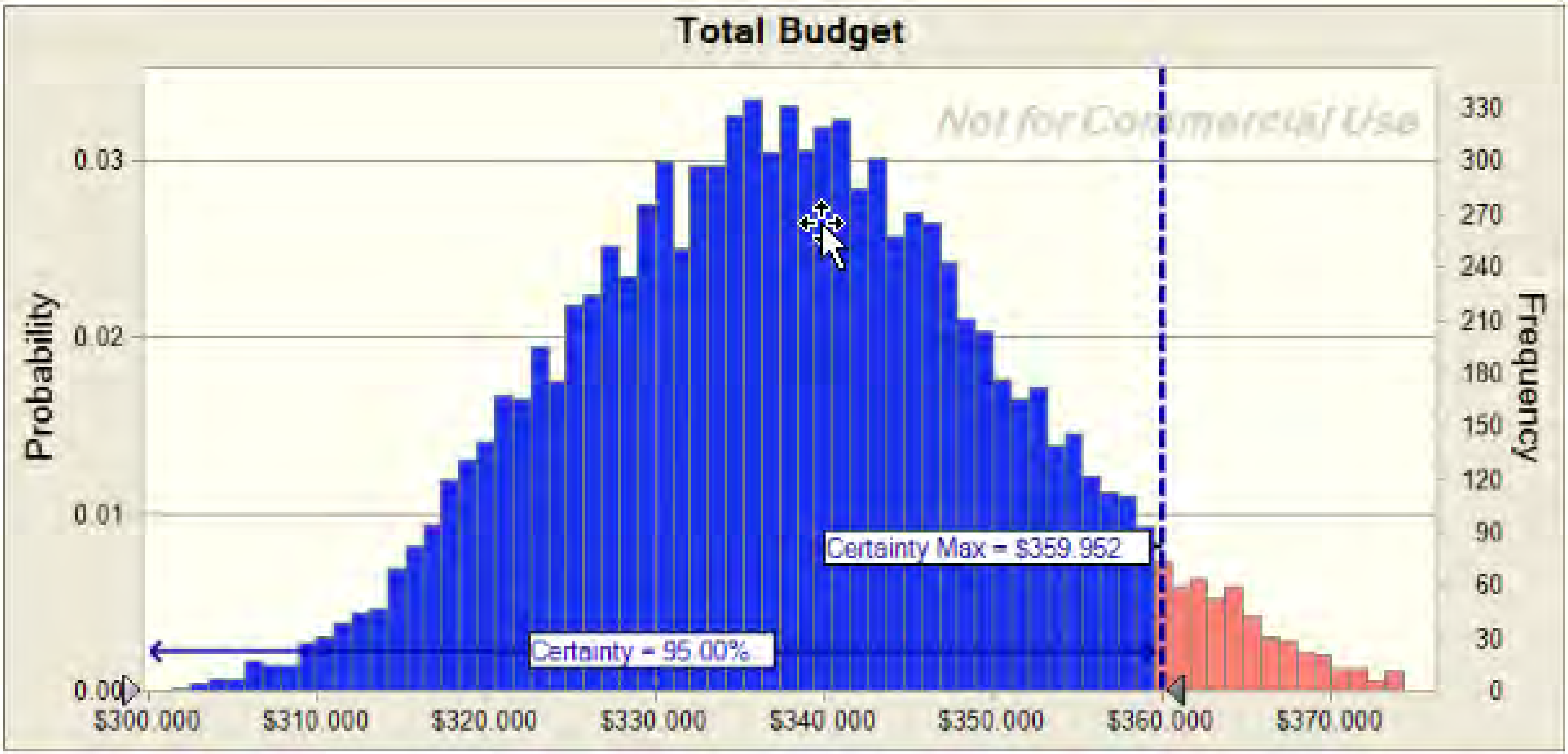
After 795 solutions were evaluated in 10 minutes and 1 second, the 5% Percentile of Total Customer Value was improved to 20.54

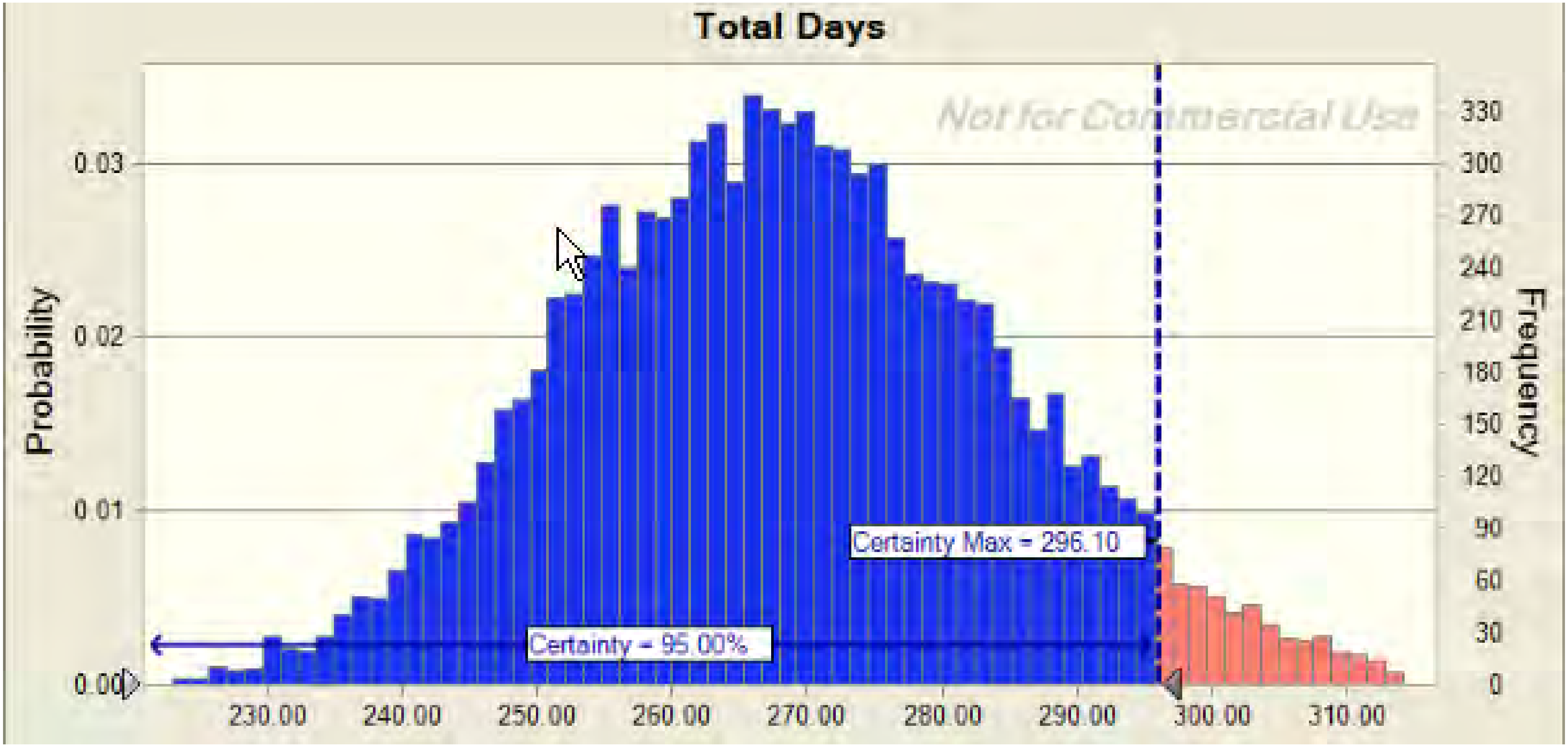


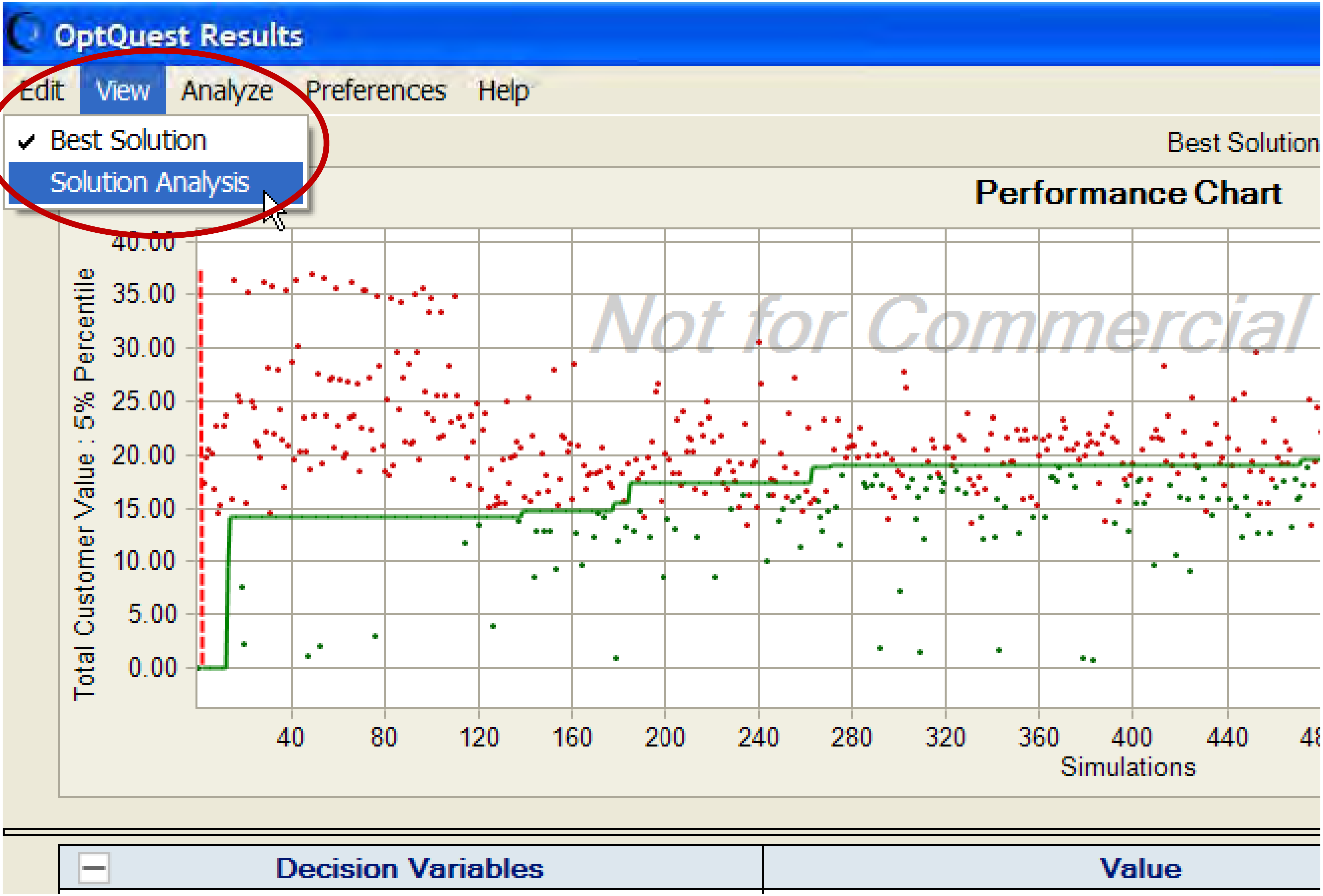

| | | |
|---|---|---|
| **Objectives** | Best Solution: | |
| Maximize the 5% Percentile of Total Customer Value | 20.54 | Cell: AG26 |
| **Requirements** | | |
| The 95% Percentile of Total Senior Resource must be less than 2.10 | 1.57 | Cell: AC26 |
| The 95% Percentile of Total Days must be less than 300.00 | 296.10 | Cell: Y26 |
| The 95% Percentile of Total Budget must be less than $400.000 | $359.952 | Cell: O26 |




| Statistics: | Forecast values |
|---|---|
| Trials | 10,000 |
| Mean | 1.57 |
| Median | 1.57 |
| Mode | 1.57 |


Total Budget
Not for Commercial Use
Probability
Frequency
0.03
0.02
0.01
0.00
330
300
270
240
210
180
150
120
90
60
30
0
Certainty Max = $359.952
Certainty = 95.00%
$300.000
$310.000
$320.000
$330.000
$340.000
$350.000
$360.000
$370.000


Total Days
Not for Commercial Use
Probability
Frequency
0.03
0.02
0.01
0.00
330
300
270
240
210
180
150
120
90
60
30
0
Certainty Max = 296.10
Certainty = 95.00%
230.00
240.00
250.00
260.00
270.00
280.00
290.00
300.00
310.00



OptQuest Results
Edit
View
Analyze
Preferences
Help
Best Solution
Solution Analysis
Best Solution
Performance Chart
Total Customer Value : 5% Percentile
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
40
80
120
160
200
240
280
320
360
400
440
Simulations
Not for Commercial
Decision Variables
Value

795 Total Solutions | Solution Analysis View | 258 Displayed

| | | | Objective | Requirements | | | Decision Va |
|---|---|---|---|---|---|---|---|
| | Rank | Solution # | Maximize 5% Percentile Total Customer Value | 95% Percentile < 2.10 Total Senior Resource | 95% Percentile < 300.00 Total Days | 95% Percentile < $400.000 Total Budget | Feature 1 Decisi |
| | 1 | 656 | 20.54 | 1.57 | 296.10 | $359.952 | 0 |
| † | 2 | 772 | 20.51 | 1.23 | 289.29 | $371.264 | 0 |
| | 3 | 489 | 19.64 | 1.65 | 295.60 | $329.433 | 0 |
| † | 4 | 740 | 19.64 | 1.48 | 296.89 | $286.605 | 0 |
| | 5 | 472 | 19.59 | 1.57 | 287.43 | $325.140 | 1 |
| † | 6 | 627 | 19.39 | 1.65 | 293.52 | $299.642 | 0 |
| † | 7 | 768 | 19.36 | 1.48 | 283.53 | $345.677 | 0 |
| † | 8 | 549 | 19.31 | 1.19 | 282.13 | $292.331 | 0 |
| | 9 | 284 | 19.09 | 1.70 | 294.06 | $286.844 | 1 |
| | 10 | 271 | 19.04 | 1.36 | 283.31 | $316.881 | 0 |
| † | 11 | 673 | 18.91 | 0.83 | 279.33 | $369.156 | 0 |
| † | 12 | 368 | 18.89 | 1.82 | 286.77 | $291.631 | 0 |
| † | 13 | 428 | 18.84 | 1.06 | 272.43 | $316.042 | 1 |
| † | 14 | 494 | 18.84 | 1.40 | 270.25 | $301.929 | 0 |
| | 15 | 263 | 18.79 | 1.55 | 296.84 | $293.061 | 1 |
| † | 16 | 474 | 18.79 | 1.67 | 297.46 | $328.196 | 1 |
| † | | | | | | | |

Statistics: † - Low-confidence solution (values are approx.)

| | | | | | | |
|---|---|---|---|---|---|---|
| Minimum | | 0.00 | 0.00 | 0.00 | $0.000 | 0.00 |
| Mean | | 13.98 | 1.06 | 238.17 | $248.456 | 0.35 |
| Maximum | | 20.54 | 2.06 | 299.12 | $371.264 | 1.00 |
| Std. Dev. | | 4.18 | 0.45 | 59.66 | $70.477 | 0.48 |

Show the best

- ○ 15 solutions
- ○ 10 % of solutions
- ◉ All feasible solutions (258)
- ○ New best solutions (12)

Include

- ☑ Feasible solutions (258)
- ☐ Infeasible solutions (537)

| Objectives | Best Solution: | |
|---|---|---|
| Maximize the 5% Percentile of Total Customer Value | 20.01 | Cell: AG26 |
| **Requirements** | | |
| The 95% Percentile of Total Senior Resource must be less than 1.00 | 0.41 | Cell: AC26 |
| The 95% Percentile of Total Days must be less than 300.00 | 293.86 | Cell: Y26 |
| The 95% Percentile of Total Budget must be less than $400.000 | $351.343 | Cell: O26 |

| Decision variables | | Best Solution: | |
|---|---|---|---|
| Feature 1 Decision | | 0.00 | Cell: A5 |
| Feature 10 Decision | ✓ | 1.00 | Cell: A14 |
| Feature 11 Decision | ✓ | 1.00 | Cell: A15 |
| Feature 12 Decision | | 0.00 | Cell: A16 |
| Feature 13 Decision | | 0.00 | Cell: A17 |
| Feature 14 Decision | | 0.00 | Cell: A18 |
| Feature 15 Decision | ✓ | 1.00 | Cell: A19 |
| Feature 16 Decision | | 0.00 | Cell: A20 |
| Feature 17 Decision | ✓ | 1.00 | Cell: A21 |
| Feature 18 Decision | ✓ | 1.00 | Cell: A22 |
| Feature 19 Decision | ✓ | 1.00 | Cell: A23 |
| Feature 2 Decision | ✓ | 1.00 | Cell: A6 |
| Feature 20 Decision | ✓ | 1.00 | Cell: A24 |
| Feature 3 Decision | | 0.00 | Cell: A7 |
| Feature 4 Decision | | 0.00 | Cell: A8 |
| Feature 5 Decision | | 0.00 | Cell: A9 |
| Feature 6 Decision | | 0.00 | Cell: A10 |
| Feature 7 Decision | | 0.00 | Cell: A11 |
| Feature 8 Decision | | 0.00 | Cell: A12 |
| Feature 9 Decision | ✓ | 1.00 | Cell: A13 |

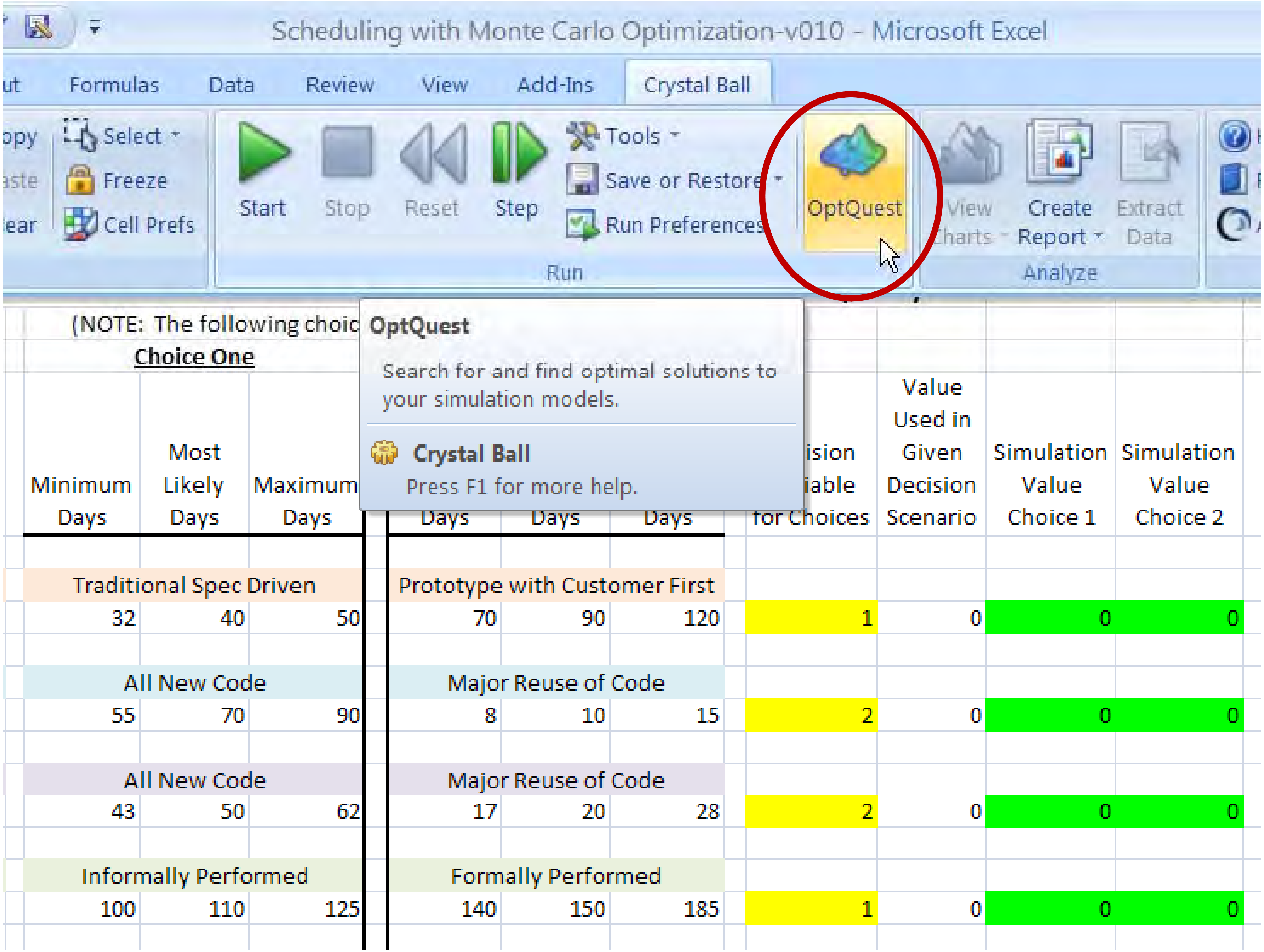
# PPM Exercise 2: Scheduling Projects with Monte Carlo Simulation and Optimization

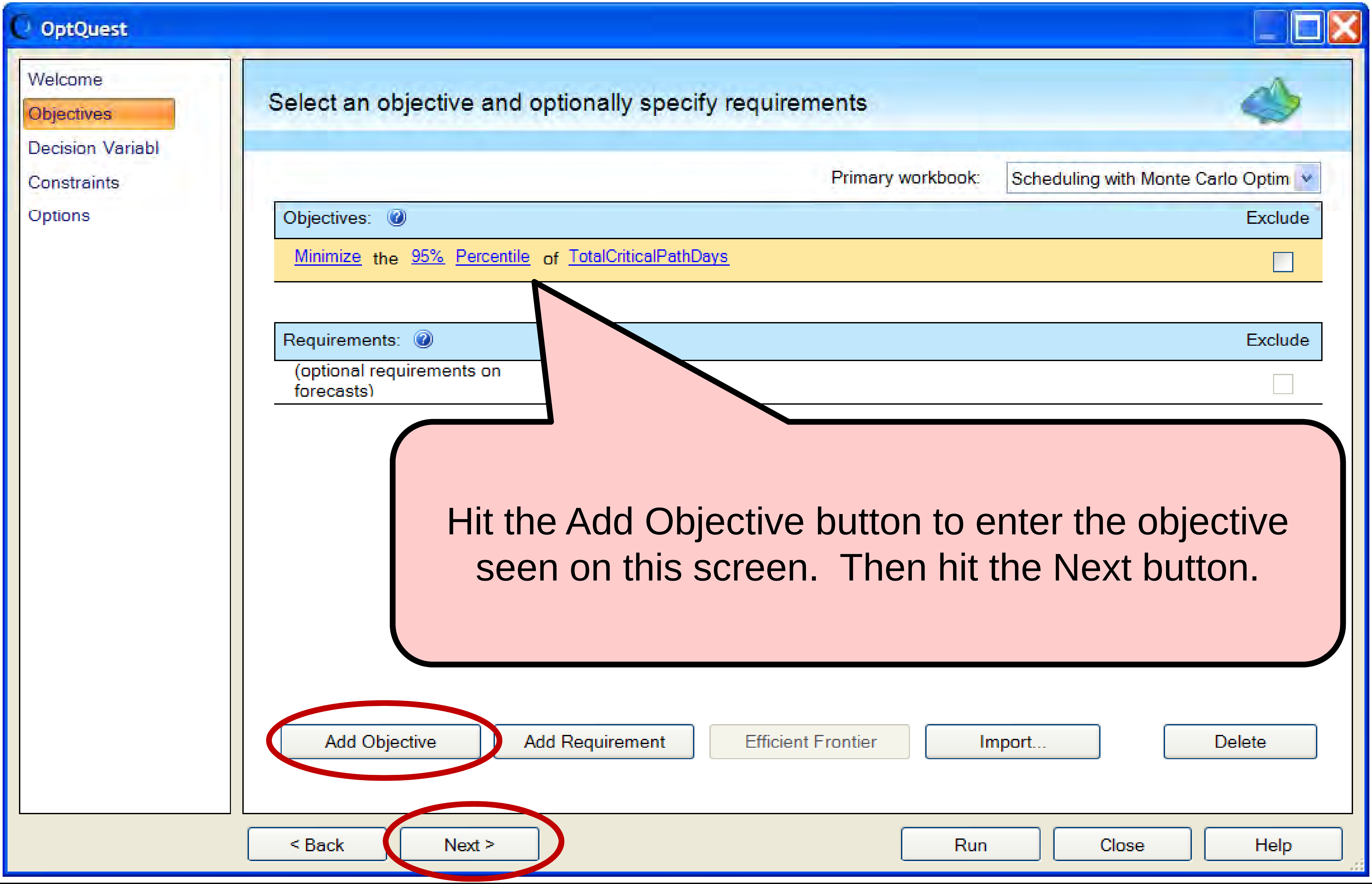
# Scheduling With Monte Carlo Optimization-v010.xls file

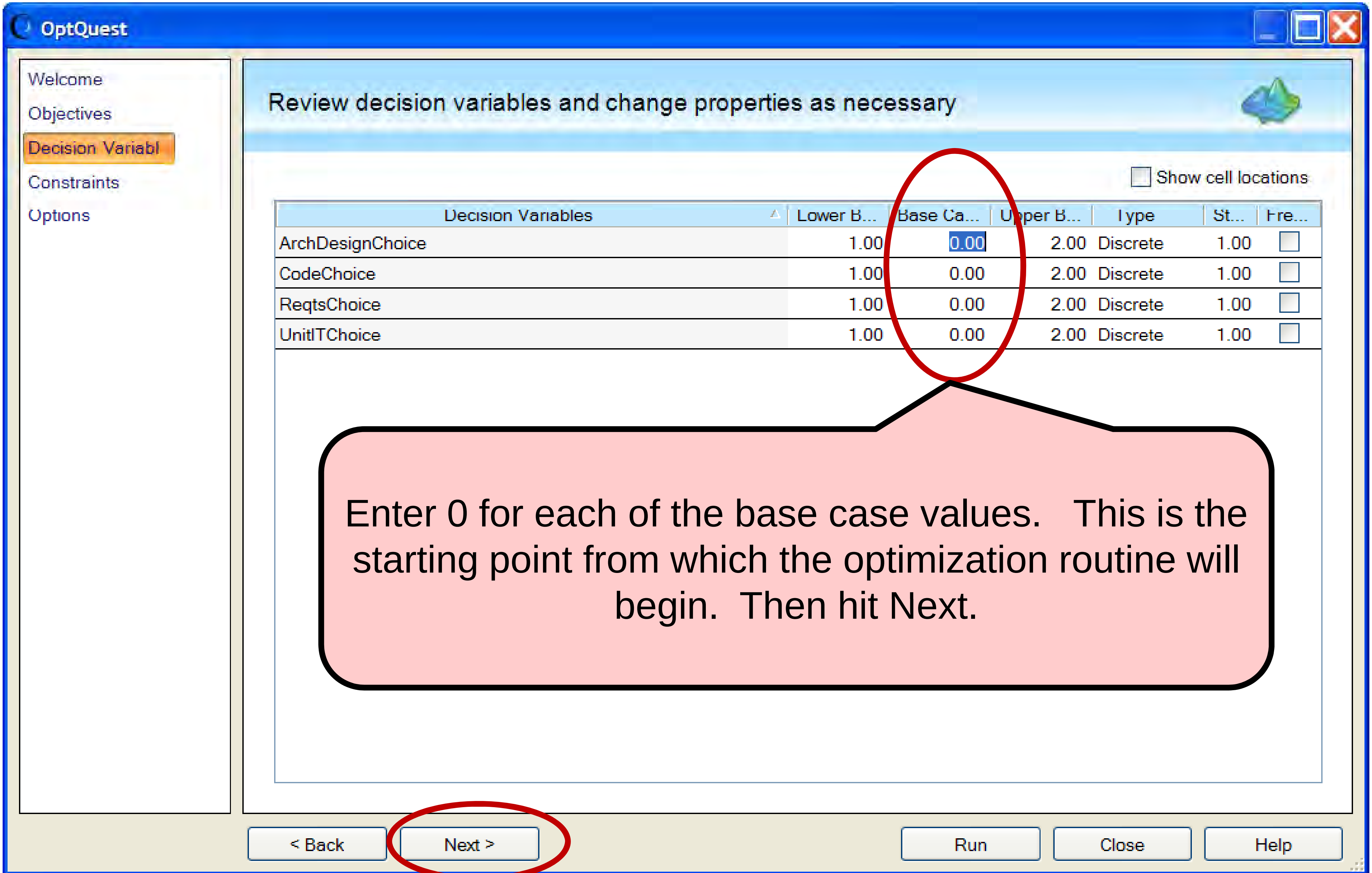
**Example of Using Monte Carlo Simulation and Optimization to Make Decisions**
**Regarding the Work Activities and Schedule for the Hospital Records Information System (HRIS)**

(NOTE: The following choices are independent decisions)

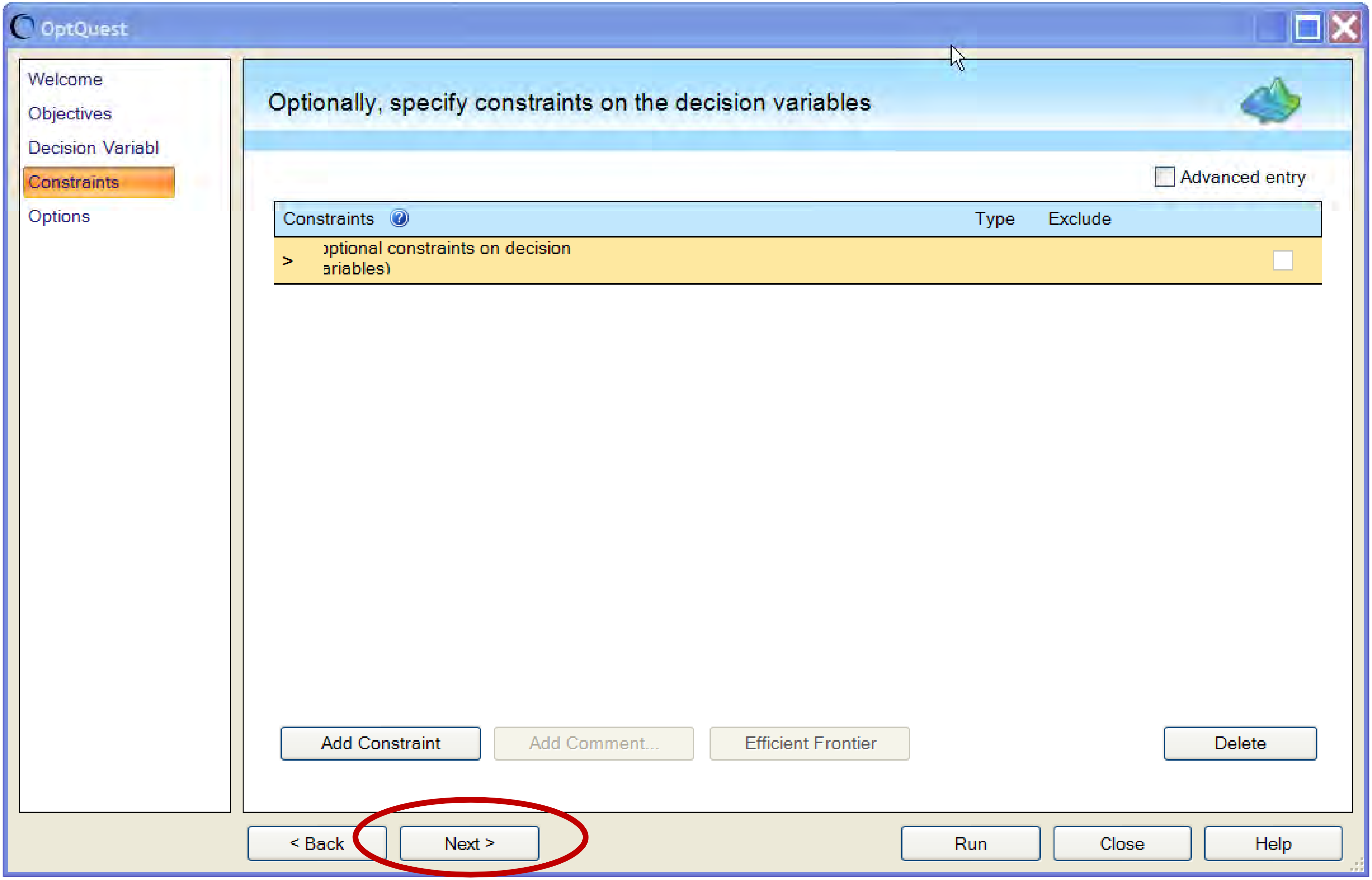
| | | Choice One | | | Choice Two | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Critical Path Tasks** | **Task Description** | Minimum Days | Most Likely Days | Maximum Days | Minimum Days | Most Likely Days | Maximum Days | Decision Variable for Choices | Value Used in Given Decision Scenario | Simulation Value Choice 1 | Simulation Value Choice 2 |
| 1 | Reqts Development | Traditional Spec Driven | | | Prototype with Customer First | | | | | | |
| | | 32 | 40 | 50 | 70 | 90 | 120 | 1 | 0 | 0 | 0 |
| 2 | Architecture/Design | All New Code | | | Major Reuse of Code | | | | | | |
| | | 55 | 70 | 90 | 8 | 10 | 15 | 2 | 0 | 0 | 0 |
| 3 | Code | All New Code | | | Major Reuse of Code | | | | | | |
| | | 43 | 50 | 62 | 17 | 20 | 28 | 2 | 0 | 0 | 0 |
| 4 | Unit/Integration Test | Informally Performed | | | Formally Performed | | | | | | |
| | | 100 | 110 | 125 | 140 | 150 | 185 | 1 | 0 | 0 | 0 |
| 5 | Acceptance Test | Only Choice | | | | | | | | | |
| | | 19 | 30 | 39 | | | | 1 | 0 | 0 | |
| | | | | | | | | **Total Critical Path Days >>** | **0** | | |

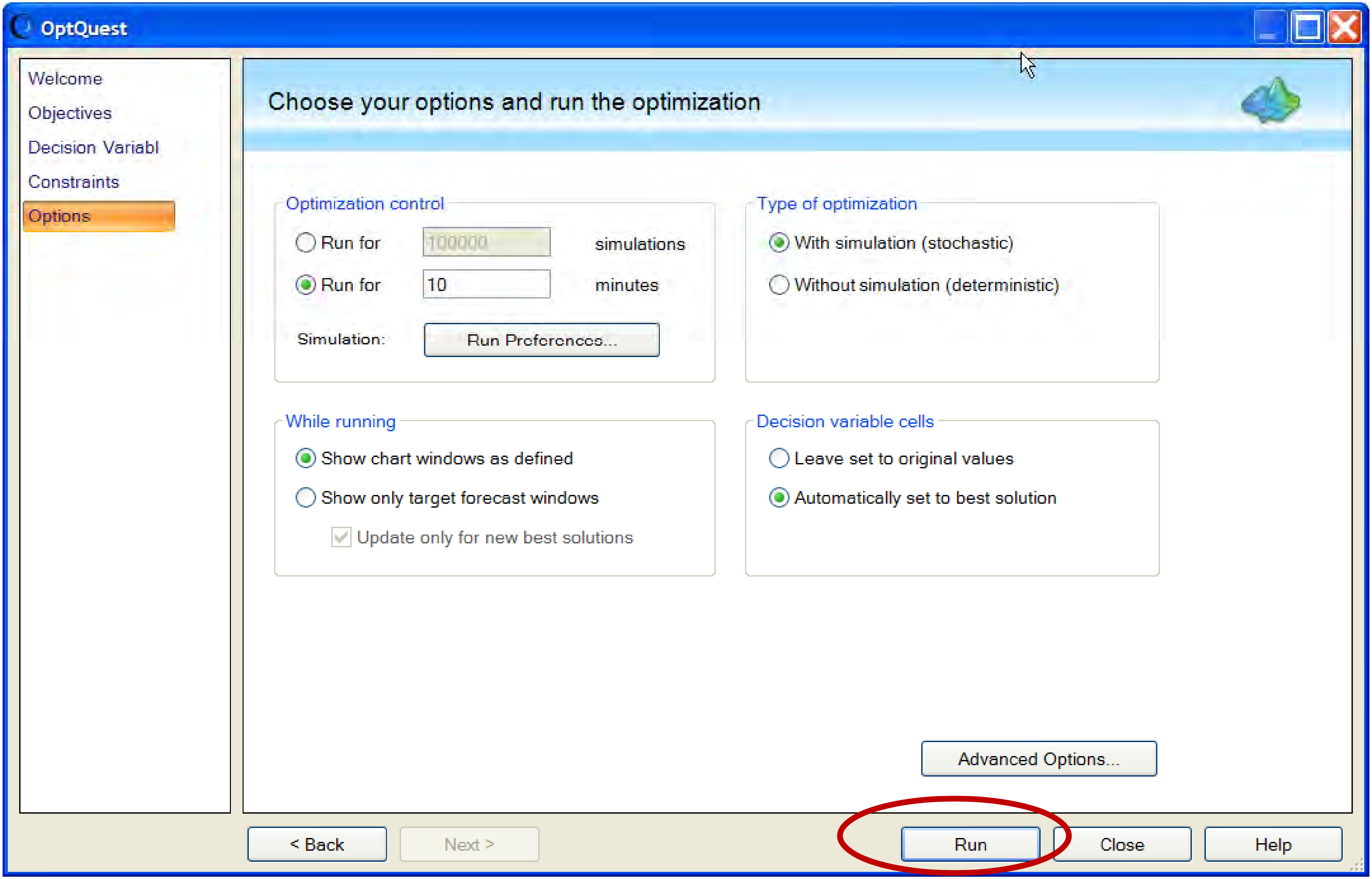

Scheduling with Monte Carlo Optimization-v010 - Microsoft Excel
Formulas
Data
Review
View
Add-Ins
Crystal Ball
Select
Freeze
Cell Prefs
Start
Stop
Reset
Step
Tools
Save or Restore
Run Preferences
Run
OptQuest
View Charts
Create Report
Extract Data
Analyze
(NOTE: The following choic
Choice One
OptQuest
Search for and find optimal solutions to your simulation models.
Crystal Ball
Press F1 for more help.
Minimum Days
Most Likely Days
Maximum Days
Days
Days
Days
for Choices
Value Used in Given Decision Scenario
Simulation Value Choice 1
Simulation Value Choice 2
Traditional Spec Driven
Prototype with Customer First
32 40 50 70 90 120 1 0 0 0
All New Code
Major Reuse of Code
55 70 90 8 10 15 2 0 0 0
All New Code
Major Reuse of Code
43 50 62 17 20 28 2 0 0 0
Informally Performed
Formally Performed
100 110 125 140 150 185 1 0 0 0

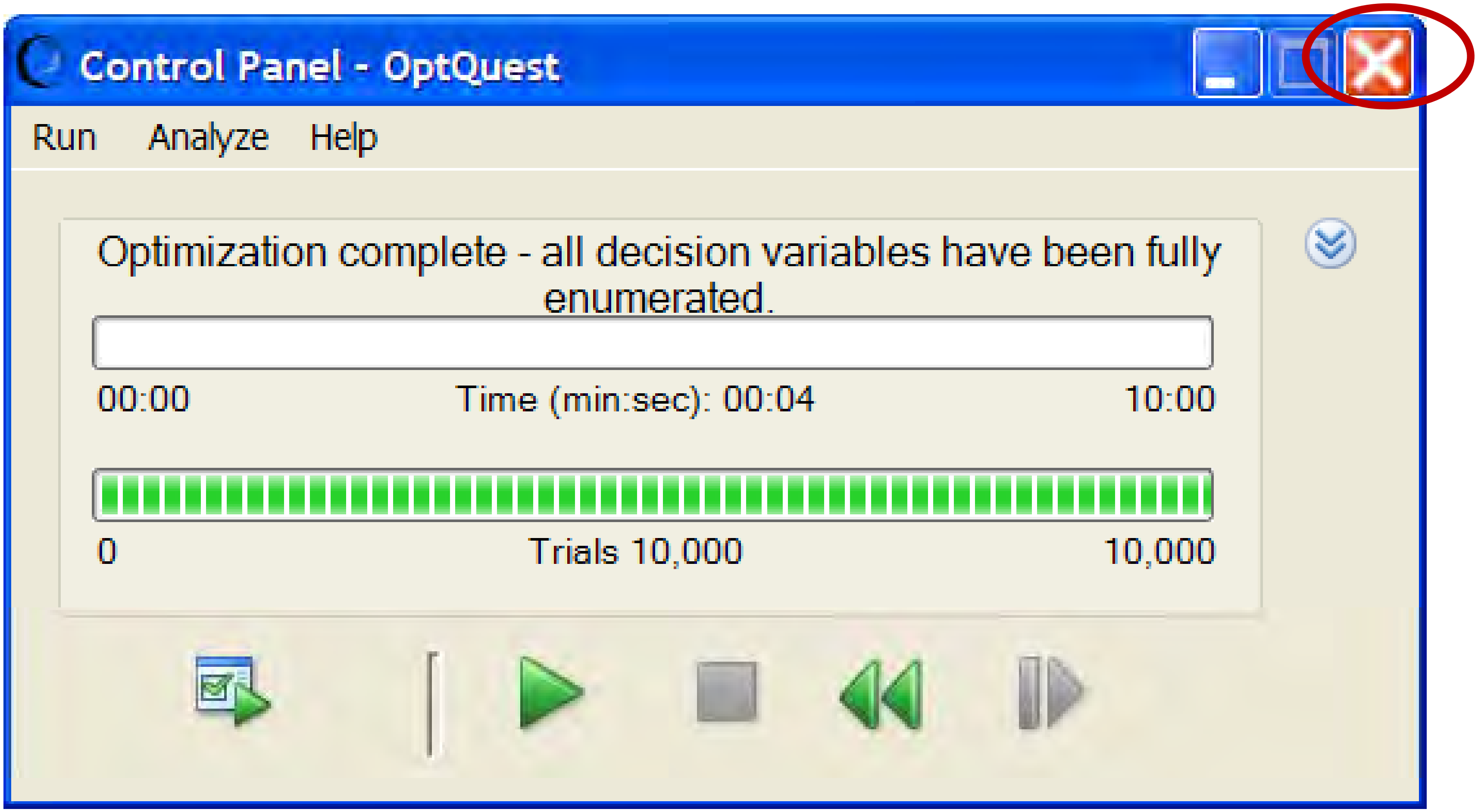

OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Select an objective and optionally specify requirements
Primary workbook:
Scheduling with Monte Carlo Optim
Objectives:
Exclude
Minimize the 95% Percentile of TotalCriticalPathDays
Requirements:
Exclude
(optional requirements on forecasts)
Hit the Add Objective button to enter the objective seen on this screen. Then hit the Next button.
Add Objective
Add Requirement
Efficient Frontier
Import...
Delete
< Back
Next >
Run
Close
Help

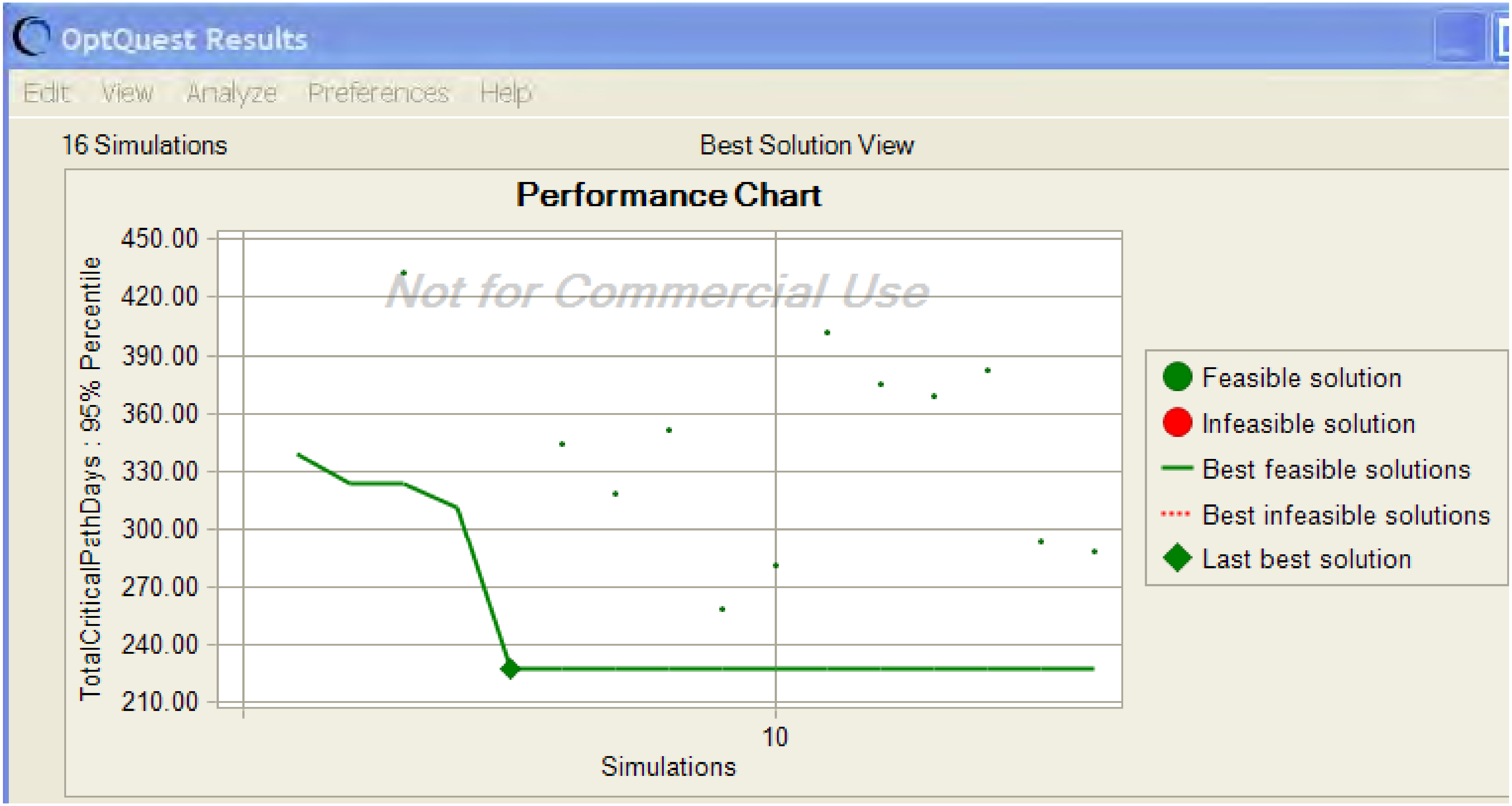



Enter 0 for each of the base case values. This is the starting point from which the optimization routine will begin. Then hit Next.


OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Optionally, specify constraints on the decision variables
Advanced entry
Constraints
Type
Exclude
>
ɔptional constraints on decision
ariables)
Add Constraint
Add Comment...
Efficient Frontier
Delete
< Back
Next >
Run
Close
Help


OptQuest
Welcome
Objectives
Decision Variabl
Constraints
Options
Choose your options and run the optimization
Optimization control
Run for
100000
simulations
Run for
10
minutes
Simulation:
Run Preferences...
Type of optimization
With simulation (stochastic)
Without simulation (deterministic)
While running
Show chart windows as defined
Show only target forecast windows
Update only for new best solutions
Decision variable cells
Leave set to original values
Automatically set to best solution
Advanced Options...
< Back
Next >
Run
Close
Help


Control Panel - OptQuest
Run
Analyze
Help
Optimization complete - all decision variables have been fully enumerated.
00:00
Time (min:sec): 00:04
10:00
0
Trials 10,000
10,000


OptQuest Results
Edit View Analyze Preferences Help
16 Simulations
Best Solution View
Performance Chart
Not for Commercial Use
TotalCriticalPathDays : 95% Percentile
450.00
420.00
390.00
360.00
330.00
300.00
270.00
240.00
210.00
10
Simulations
Feasible solution
Infeasible solution
Best feasible solutions
Best infeasible solutions
Last best solution

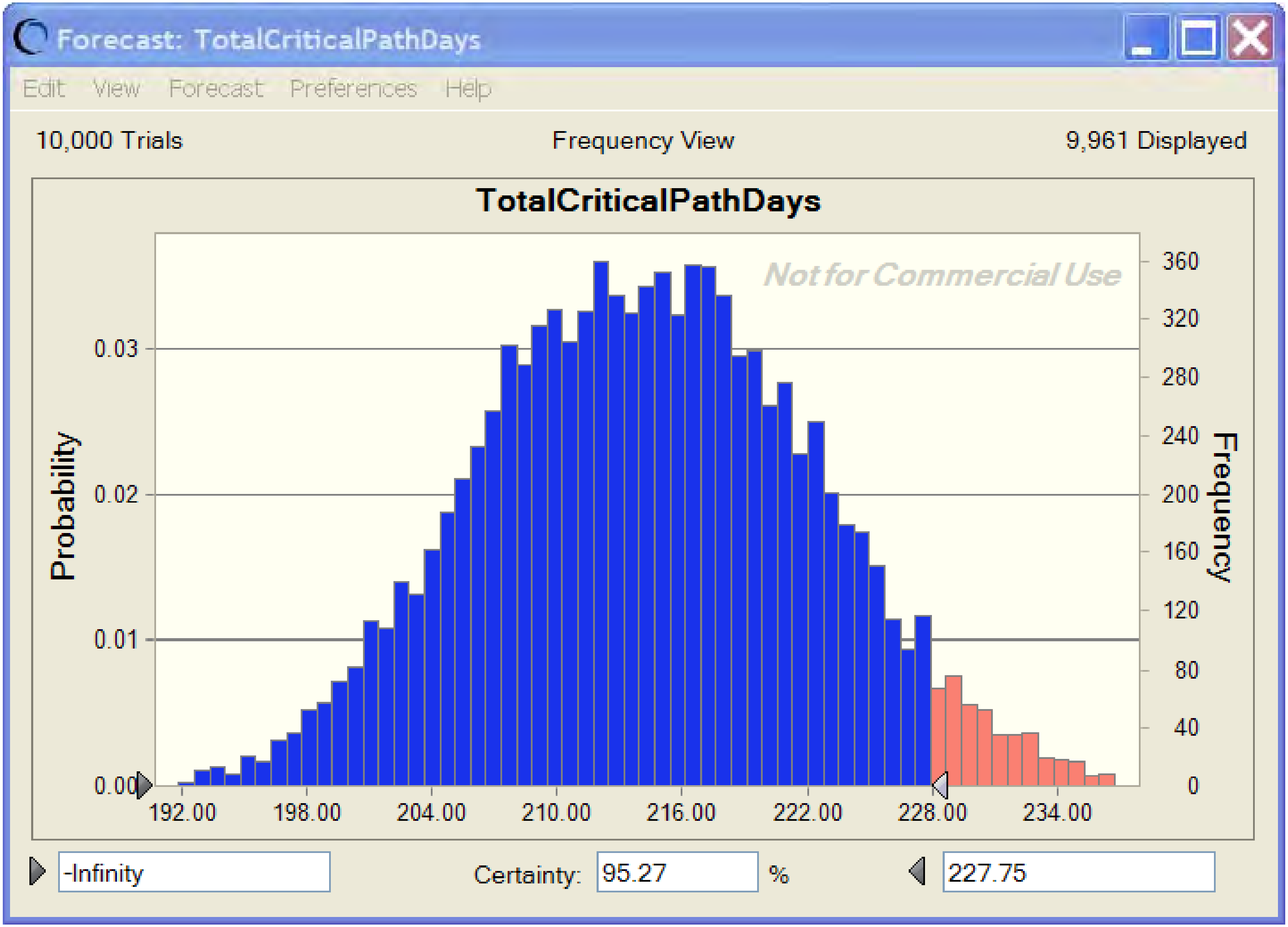
Best Solution: Simulation # 5

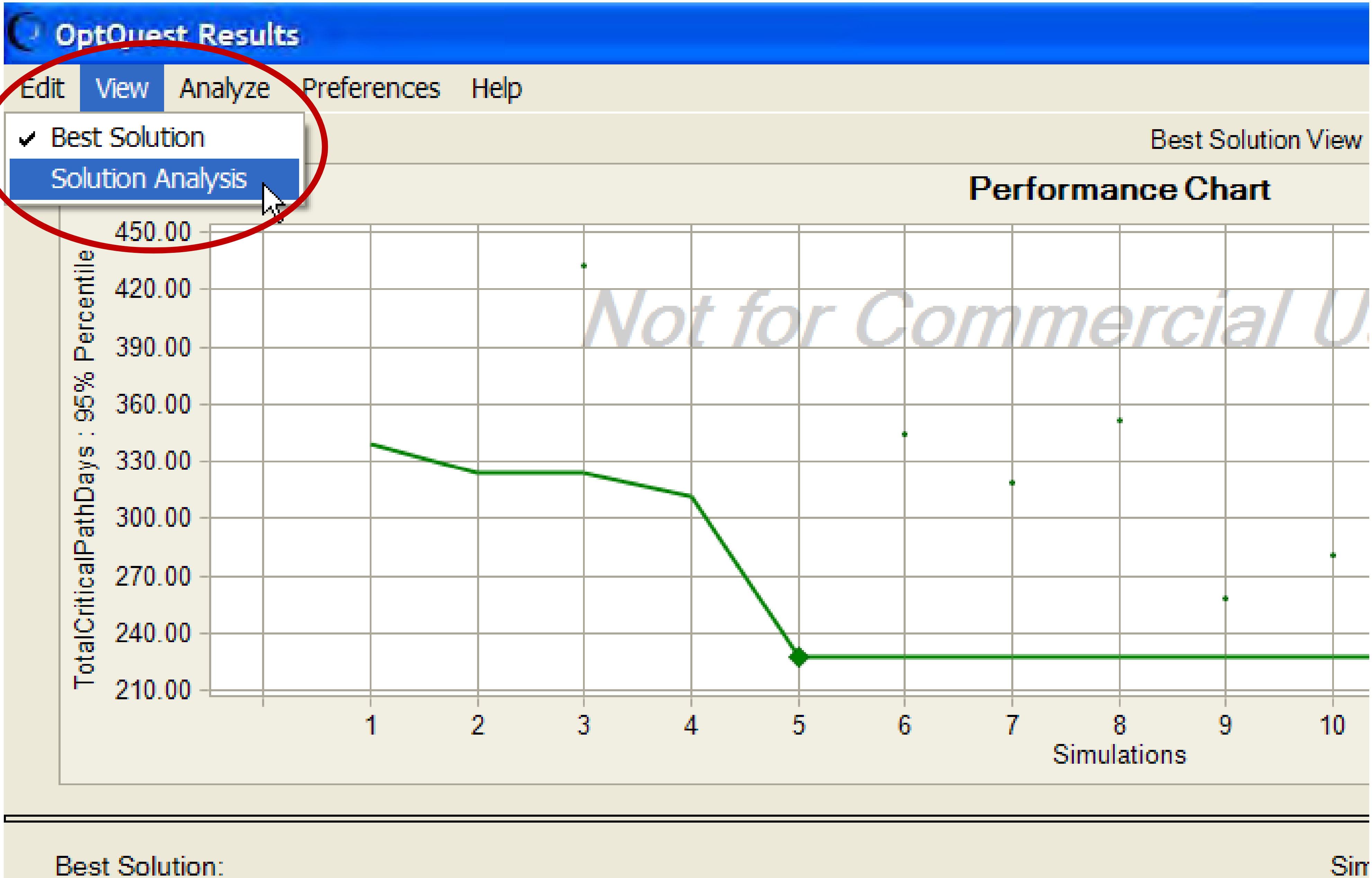
| Objectives | Value |
|---|---|
| Minimize the 95% Percentile of TotalCriticalPathDays | 227.59 |

| Requirements | Value |
|---|---|

| Constraints | Left Side | | Right Side |
|---|---|---|---|

| Decision Variables | Value |
|---|---|
| ArchDesignChoice | 2.00 |
| CodeChoice | 2.00 |
| ReqtsChoice | 1.00 |
| UnitITChoice | 1.00 |


Forecast: TotalCriticalPathDays
Edit View Forecast Preferences Help
10,000 Trials
Frequency View
9,961 Displayed
TotalCriticalPathDays
Not for Commercial Use
Probability
Frequency
0.03
0.02
0.01
0.00
360
320
280
240
200
160
120
80
40
0
192.00
198.00
204.00
210.00
216.00
222.00
228.00
234.00
-Infinity
Certainty: 95.27 %
227.75


OptQuest Results
Edit
View
Analyze
Preferences
Help
Best Solution
Solution Analysis
Best Solution View
Performance Chart
Not for Commercial U
TotalCriticalPathDays : 95% Percentile
450.00
420.00
390.00
360.00
330.00
300.00
270.00
240.00
210.00
1
2
3
4
5
6
7
8
9
10
Simulations
Best Solution:
Sim

16 Total Solutions

Solution Analysis View

| | | Objective | Decision Variables | | | |
|---|---|---|---|---|---|---|
| Rank | Solution # | Minimize 95% Percentile TotalCriticalPathDays | ArchDesignChoice | CodeChoice | ReqtsChoice | UnitITChoice |
| 1 | 5 | 227.59 | 2.00 | 2.00 | 1.00 | 1.00 |
| 2 | 9 | 258.40 | 2.00 | 1.00 | 1.00 | 1.00 |
| 3 | 10 | 280.93 | 2.00 | 2.00 | 1.00 | 2.00 |
| 4 | 16 | 288.51 | 2.00 | 2.00 | 2.00 | 1.00 |
| 5 | 15 | 293.13 | 1.00 | 2.00 | 1.00 | 1.00 |
| 6 | 4 | 311.57 | 2.00 | 1.00 | 1.00 | 2.00 |
| 7 | 7 | 318.81 | 2.00 | 1.00 | 2.00 | 1.00 |
| 8 | 2 | 323.84 | 1.00 | 1.00 | 1.00 | 1.00 |
| 9 | 1 | 339.05 | 2.00 | 2.00 | 2.00 | 2.00 |
| 10 | 6 | 344.50 | 1.00 | 2.00 | 1.00 | 2.00 |
| 11 | 8 | 352.16 | 1.00 | 2.00 | 2.00 | 1.00 |
| 12 | 13 | 369.57 | 2.00 | 1.00 | 2.00 | 2.00 |
| 13 | 12 | 374.86 | 1.00 | 1.00 | 1.00 | 2.00 |
| 14 | 14 | 382.32 | 1.00 | 1.00 | 2.00 | 1.00 |
| 15 | 11 | 402.05 | 1.00 | 2.00 | 2.00 | 2.00 |
| 16 | 3 | 432.65 | 1.00 | 1.00 | 2.00 | 2.00 |

# PPM Exercise 3: Predicting Product Requirements Change with Linear Regression

# Statistical Regression Landscape

The purpose of regression is to perform the basic task of ANOVA by determining whether there is significant prediction of dependent (y) variable(s) using knowledge of independent (x) variable(s).

- **Example:** Can the defects by release (y) be predicted using knowledge of one or more independent variables (x)s?
- **Some types of regression** (all y's & x's continuous unless noted as discrete):

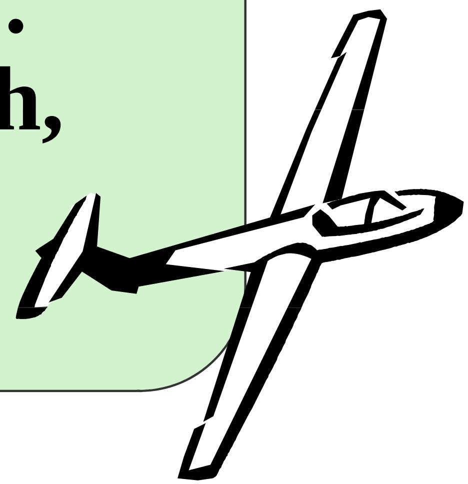
| | |
|---|---|
| Simple linear | 1 "y" & 1 "x" |
| Multiple linear | 1 "y" & multiple "x"s |
| Multivariate | multiple "y"s & 1+ "x" |
| Nonlinear | nonlinear version of the above types |
| Logistic | 1 discrete "y" & 1+ "x"s |

# p value Summary

| Method | Null | Alternative | $P < 0.05$ | $P > 0.05$ |
|---|---|---|---|---|
| Hypothesis Tests | No difference exists; no associations | Two items are different; association exists | Accept alternative | Accept null |
| Tests for Normality | Data follows Normal Distribution | Data does not follow Normal Distribution | Accept alternative | Accept null |
| ANOVA | No difference of Y across levels of x | Difference of Y exists between 1+ levels of x | Accept alternative | Accept null |
| Regression | x factor does not add value to model | X factor adds value to model | Accept alternative | Accept null |
| Chi-Square | Two discrete variables are not associated | Two discrete variables are associated | Accept alternative | Accept null |
| Logistic Regression | x factor does not add value; model has no significant x's | X factor adds value to model; model has 1+ significant x's | Accept alternative | Accept null |

# Slogan to Remember p Interpretation


**"When the p is low,
the null must go…
When the p is high,
the null must fly"**



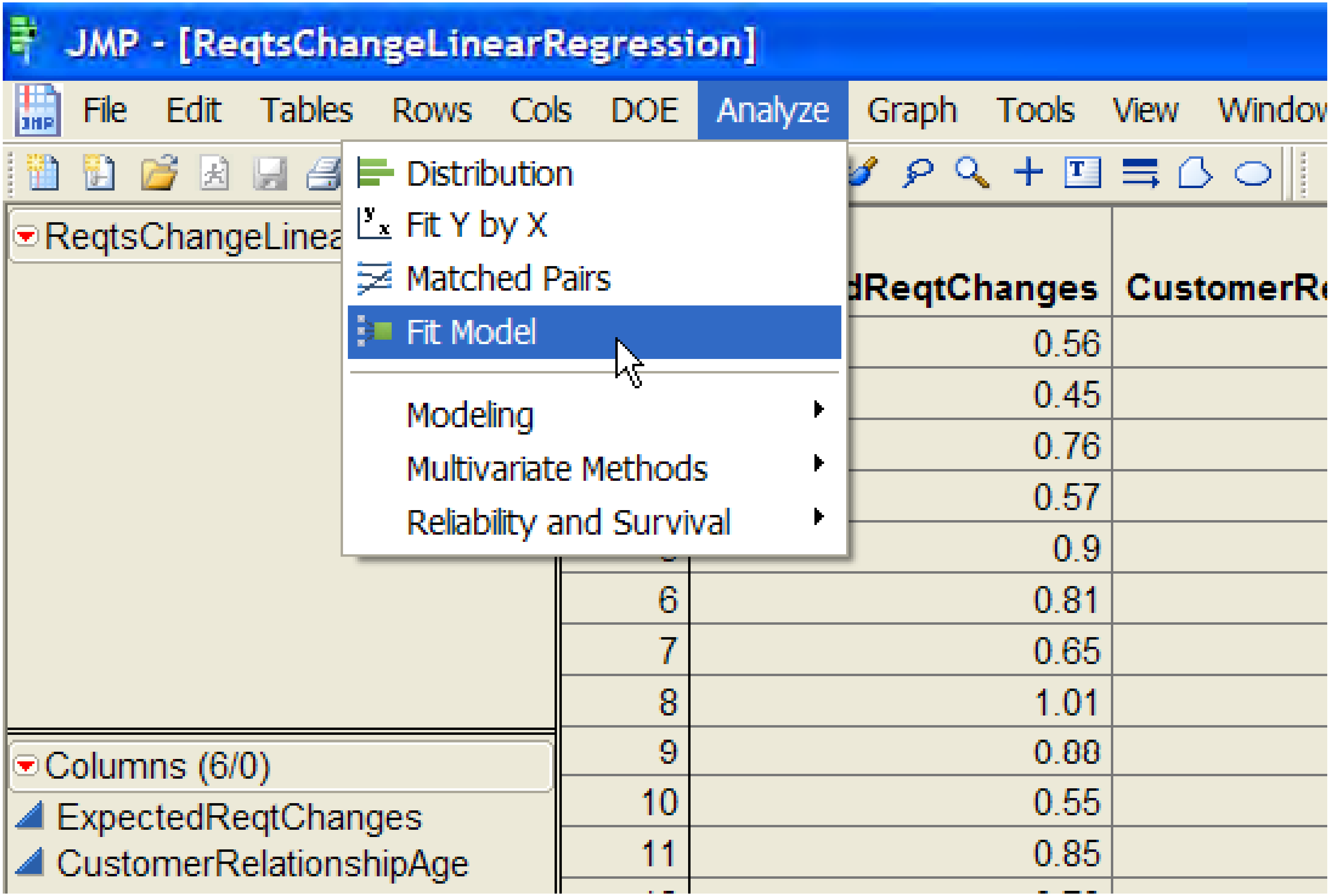
# Statistical Regression Analysis

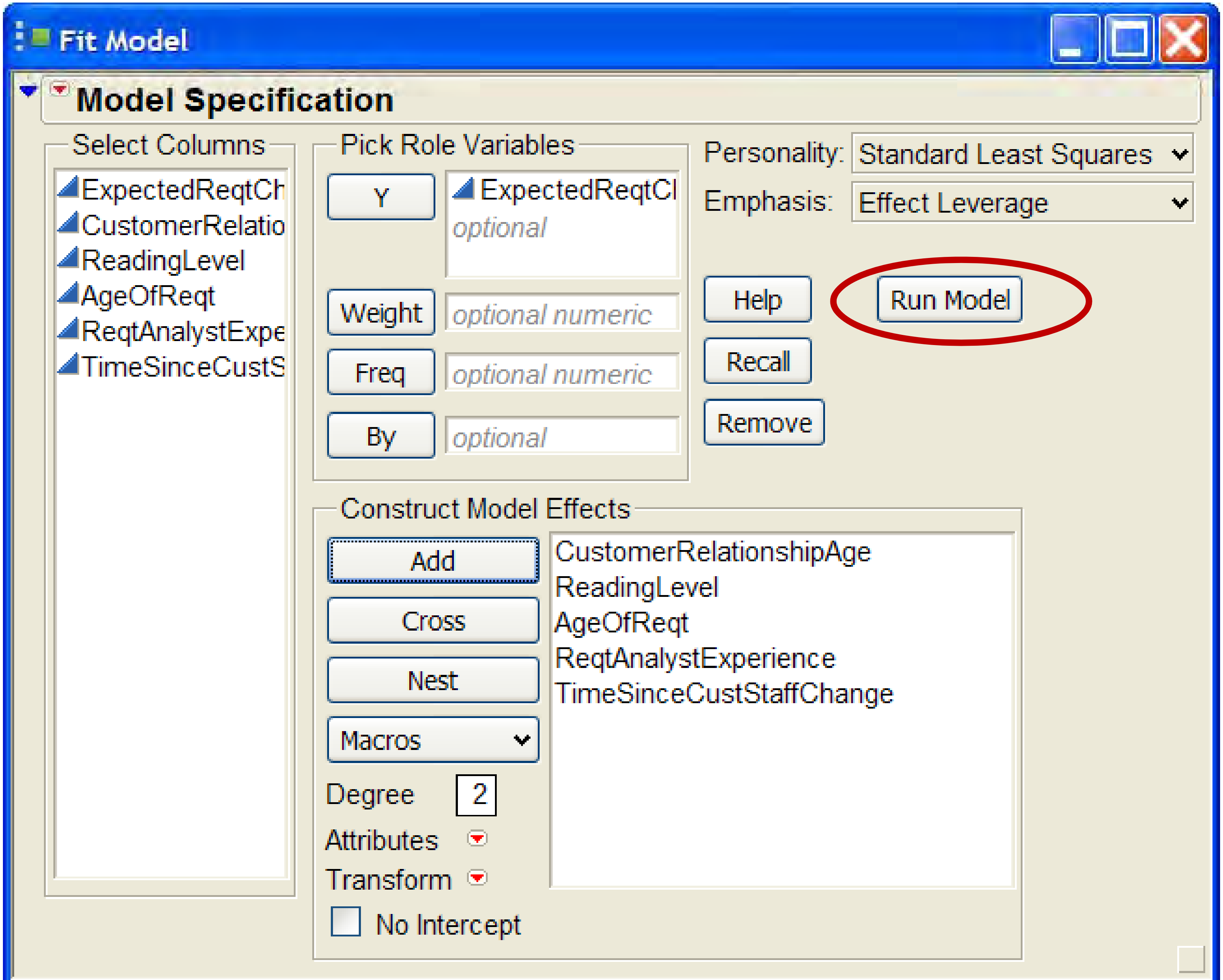
| | | Y: Continuous | Y: Discrete |
|---|---|---|---|
| X | Discrete | ANOVA and Dummy Variable Regression | Chi-Square, Logit & Logistic Regression |
| X | Continuous | Correlation & Linear Regression | Logistic Regression |

## Open the ReqtsChangeLinearRegression.jmp file

| | ExpectedReqtChanges | CustomerRelationshipAge | ReadingLevel | AgeOfReqt | ReqtAnalystExperience | TimeSinceCustStaffChange |
|---|---|---|---|---|---|---|
| 1 | 0.56 | 31.65 | 7.29 | 9.23 | 48.55 | 8.36 |
| 2 | 0.45 | 26.34 | 6.76 | 8.77 | 55.68 | 7.4 |
| 3 | 0.76 | 11.78 | 8.68 | 7.85 | 46 | 8.29 |
| 4 | 0.57 | 44.99 | 7.73 | 9.21 | 47.29 | 7.94 |
| 5 | 0.9 | 37.07 | 9.39 | 8.45 | 45.4 | 7.74 |
| 6 | 0.81 | 22.81 | 8.64 | 8.38 | 43.48 | 7.54 |
| 7 | 0.65 | 32.65 | 8.8 | 9.23 | 46.12 | 8.45 |
| 8 | 1.01 | 2.79 | 9.72 | 10.04 | 43.76 | 6.75 |
| 9 | 0.88 | 15.25 | 9.47 | 9.5 | 46.32 | 8.1 |
| 10 | 0.55 | 31.22 | 6.5 | 8.77 | 44.72 | 7.43 |
| 11 | 0.85 | 26.19 | 8.77 | 9.25 | 57.85 | 8.25 |
| 12 | 0.78 | 18.77 | 8.56 | 10.2 | 47 | 8.54 |
| 13 | 0.79 | 31.5 | 8.05 | 11.29 | 51.52 | 6.88 |
| 14 | 1.09 | 11.21 | 9.58 | 8.9 | 47.65 | 7.1 |
| 15 | 0.69 | 48.57 | 7.85 | 9.56 | 51.15 | 7.91 |
| 16 | 1.13 | 48.99 | 11.45 | 13.2 | 44.78 | 7.59 |
| 17 | 0.76 | 28.75 | 8.56 | 7.98 | 46.34 | 7.2 |
| 18 | 0.86 | 39.62 | 8.92 | 13.11 | 43.24 | 7.88 |
| 19 | 0.79 | 22.55 | 8.67 | 11.76 | 41.81 | 7.09 |
| 20 | 0.66 | 19.34 | 7.94 | 11.57 | 53.48 | 8.14 |
| 21 | 0.71 | 48.31 | 8.35 | 9.65 | 46.11 | 8.85 |
| 22 | 0.77 | 17.15 | 8.75 | 8.25 | 53.25 | 7.62 |

| Factor | Role | Data Type | Description |
|---|---|---|---|
| ExpectedReqtChanges | Y Outcome | Continuous | The number of expected changes that will occur during product development with a given product requirement |
| CustomerRelationshipAge | X1 Factor | Continuous | At the time of requirement formulation, the age in months of the relationship with the customer of the product development |
| ReadingLevel | X2 Factor | Continuous | The reading level (grade level) computed for the requirement statement (sentence or paragraph) |
| AgeOfReqt | X3 Factor | Continuous | The age in months of the product requirement at the point the requirement is identified for this product |
| ReqtAnalystExperience | X4 Factor | Continuous | The experience level in months of the Requirements Analyst |
| TimeSinceCustStaffChange | X5 Factor | Continuous | At the time of requirement formulation, the number of months since the last customer staff change |


JMP - [ReqtsChangeLinearRegression]
File
Edit
Tables
Rows
Cols
DOE
Analyze
Graph
Tools
View
Distribution
Fit Y by X
Matched Pairs
Fit Model
Modeling
Multivariate Methods
Reliability and Survival
ReqtsChangeLinea
dReqtChanges
CustomerR
0.56
0.45
0.76
0.57
0.9
6
0.81
7
0.65
8
1.01
9
0.88
10
0.55
11
0.85
Columns (6/0)
ExpectedReqtChanges
CustomerRelationshipAge


Fit Model
Model Specification
Select Columns
ExpectedReqtCh
CustomerRelatio
ReadingLevel
AgeOfReqt
ReqtAnalystExpe
TimeSinceCustS
Pick Role Variables
Y
ExpectedReqtCl
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality: Standard Least Squares
Emphasis: Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree 2
Attributes
Transform
No Intercept
CustomerRelationshipAge
ReadingLevel
AgeOfReqt
ReqtAnalystExperience
TimeSinceCustStaffChange

## Summary of Fit

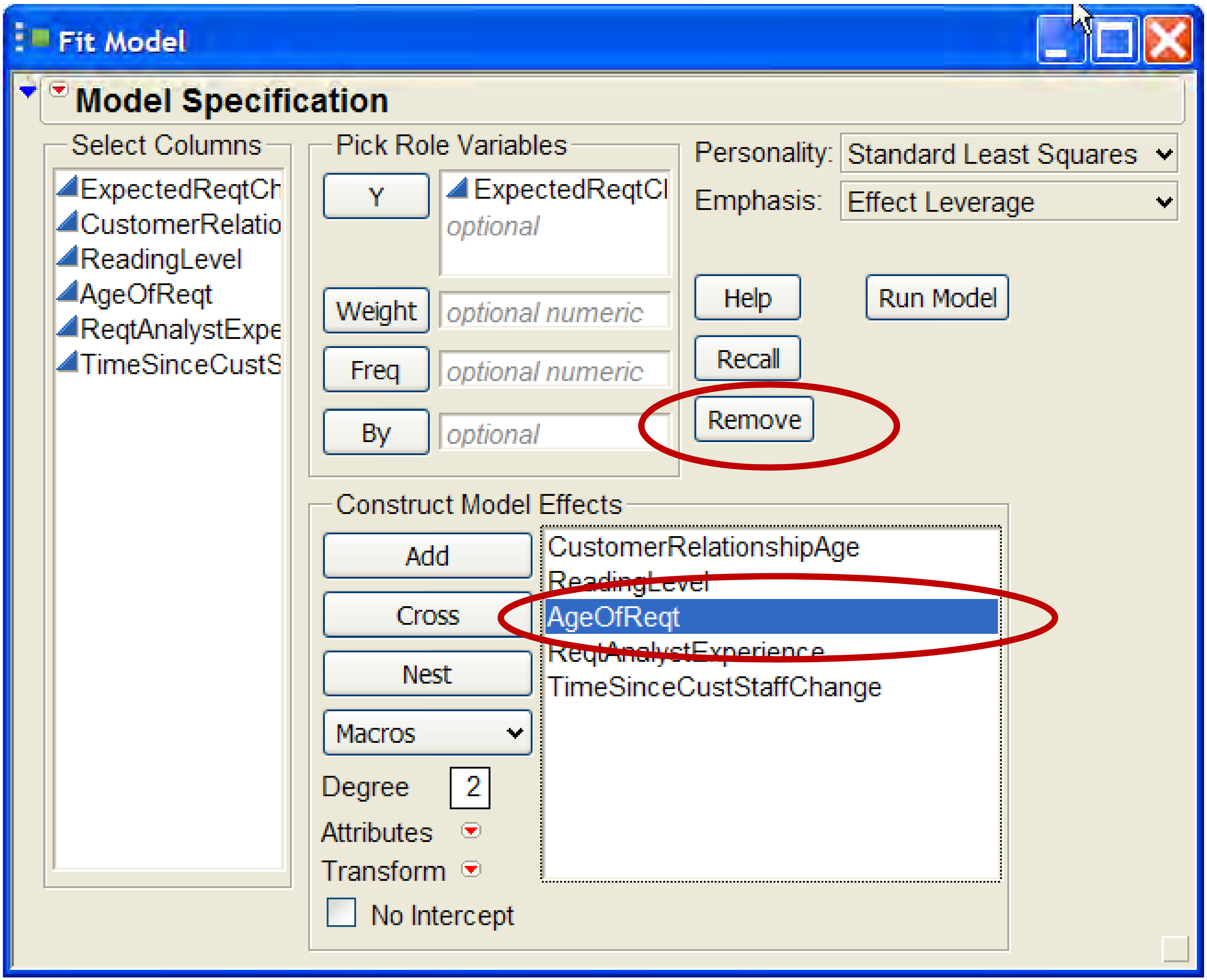
| | |
|---|---|
| RSquare | 0.903713 |
| RSquare Adj | 0.902491 |
| Root Mean Square Error | 0.063318 |
| Mean of Response | 0.844075 |
| Observations (or Sum Wgts) | 400 |

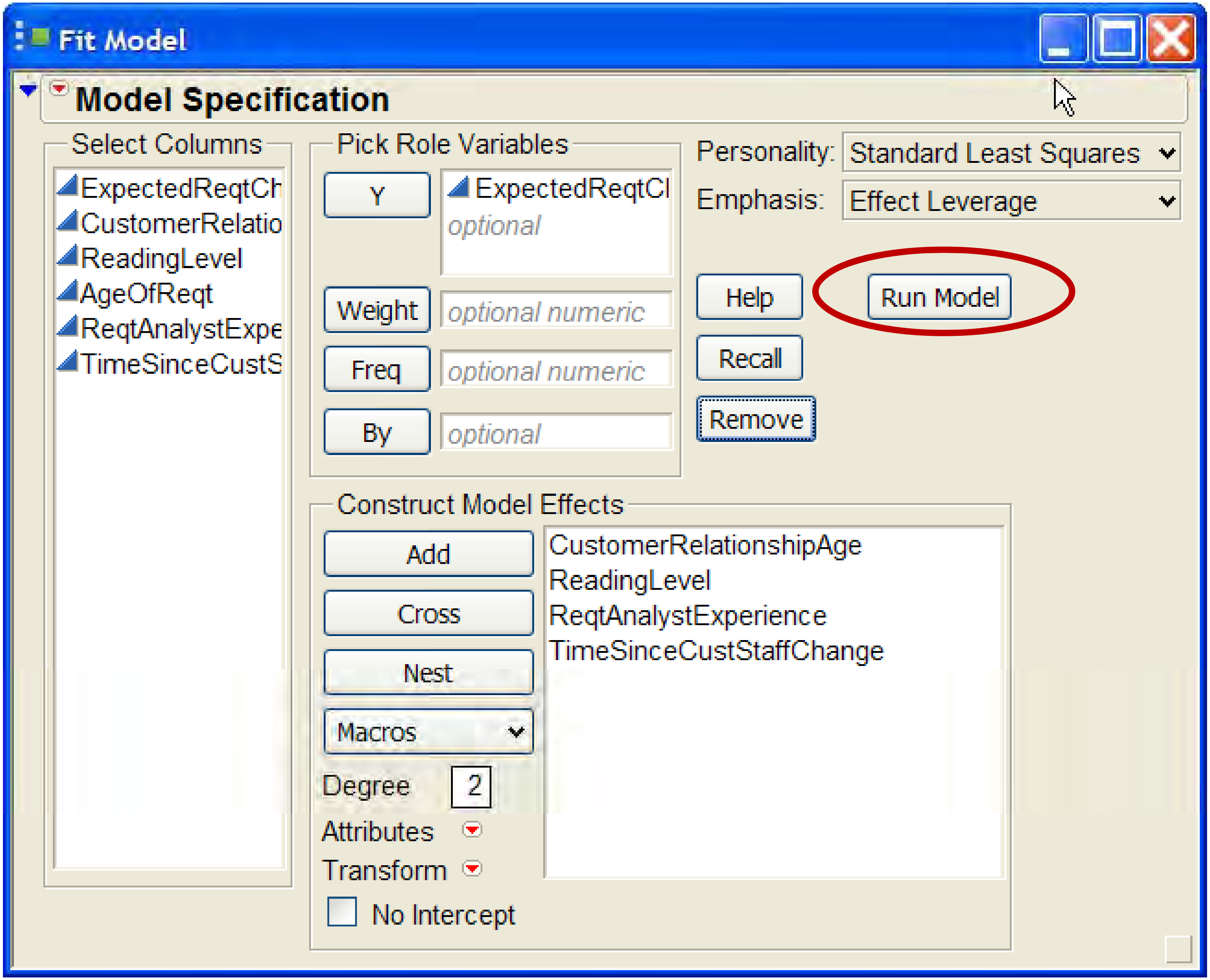
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 5 | 14.825457 | 2.96509 | 739.5833 |
| Error | 394 | 1.579600 | 0.00401 | **Prob > F** |
| C. Total | 399 | 16.405058 | | <.0001* |

## Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 0.0811043 | 0.064497 | 1.26 | 0.2093 |
| CustomerRelationshipAge | -0.000144 | 0.000262 | -0.55 | 0.5833 |
| ReadingLevel | 0.1574728 | 0.002647 | 59.49 | <.0001* |
| AgeOfReqt | 0.0004456 | 0.001195 | 0.37 | 0.7094 |
| ReqtAnalystExperience | -0.003763 | 0.000862 | -4.37 | <.0001* |
| TimeSinceCustStaffChange | -0.058817 | 0.005759 | -10.21 | <.0001* |


Fit Model
Model Specification
Select Columns
ExpectedReqtCh
CustomerRelatio
ReadingLevel
AgeOfReqt
ReqtAnalystExpe
TimeSinceCustS
Pick Role Variables
Y
ExpectedReqtCl
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
CustomerRelationshipAge
ReadingLevel
AgeOfReqt
ReqtAnalystExperience
TimeSinceCustStaffChange


Fit Model
Model Specification
Select Columns
ExpectedReqtCh
CustomerRelatio
ReadingLevel
AgeOfReqt
ReqtAnalystExpe
TimeSinceCustS
Pick Role Variables
Y
ExpectedReqtCl
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
CustomerRelationshipAge
ReadingLevel
ReqtAnalystExperience
TimeSinceCustStaffChange

## Summary of Fit

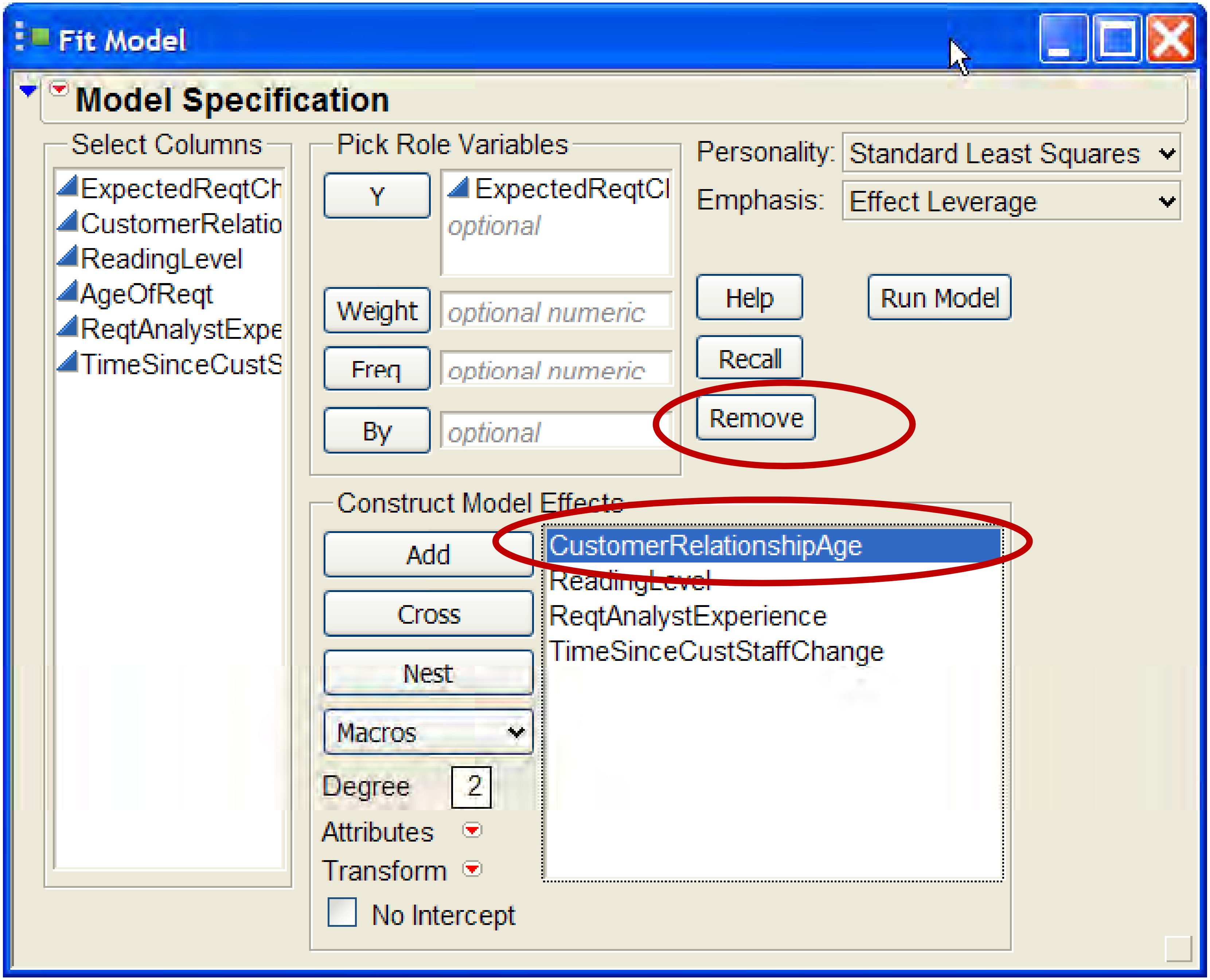
| | |
|---|---|
| RSquare | 0.903679 |
| RSquare Adj | 0.902703 |
| Root Mean Square Error | 0.063249 |
| Mean of Response | 0.844075 |
| Observations (or Sum Wgts) | 400 |

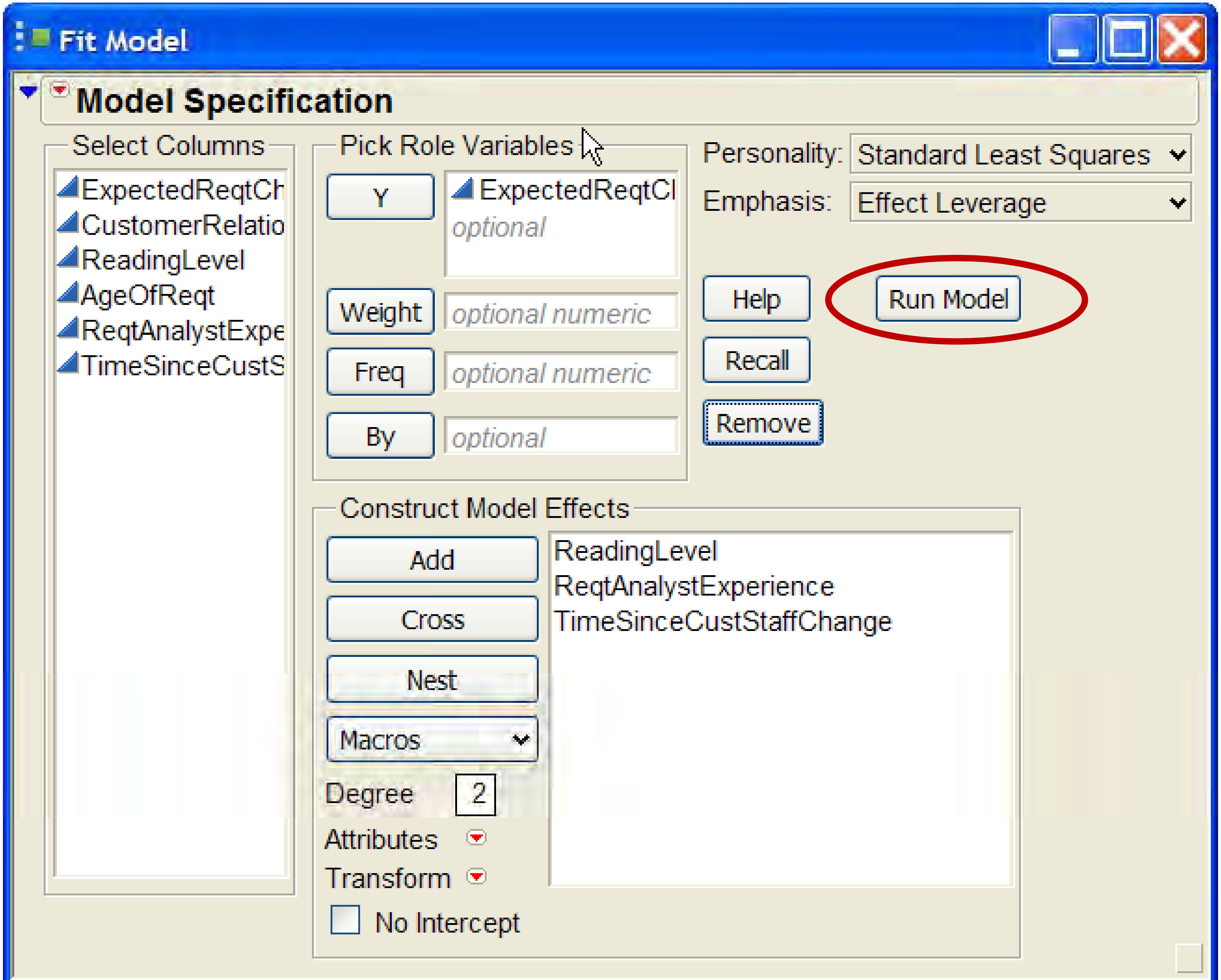
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 4 | 14.824900 | 3.70623 | 926.4638 |
| Error | 395 | 1.580158 | 0.00400 | Prob > F |
| C. Total | 399 | 16.405058 | | <.0001* |

## Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 0.0835877 | 0.064082 | 1.30 | 0.1929 |
| CustomerRelationshipAge | -0.000137 | 0.000261 | -0.53 | 0.5998 |
| ReadingLevel | 0.1575641 | 0.002633 | 59.85 | <.0001* |
| ReqtAnalystExperience | -0.003748 | 0.00086 | -4.36 | <.0001* |
| TimeSinceCustStaffChange | -0.058785 | 0.005752 | -10.22 | <.0001* |


Fit Model
Model Specification
Select Columns
ExpectedReqtCh
CustomerRelatio
ReadingLevel
AgeOfReqt
ReqtAnalystExpe
TimeSinceCustS
Pick Role Variables
Y
ExpectedReqtCl
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
CustomerRelationshipAge
ReadingLevel
Cross
ReqtAnalystExperience
Nest
TimeSinceCustStaffChange
Macros
Degree
2
Attributes
Transform
No Intercept


Fit Model
Model Specification
Select Columns
ExpectedReqtCh
CustomerRelatio
ReadingLevel
AgeOfReqt
ReqtAnalystExpe
TimeSinceCustS
Pick Role Variables
Y
ExpectedReqtCl
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality: Standard Least Squares
Emphasis: Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree 2
Attributes
Transform
No Intercept
ReadingLevel
ReqtAnalystExperience
TimeSinceCustStaffChange

## Summary of Fit

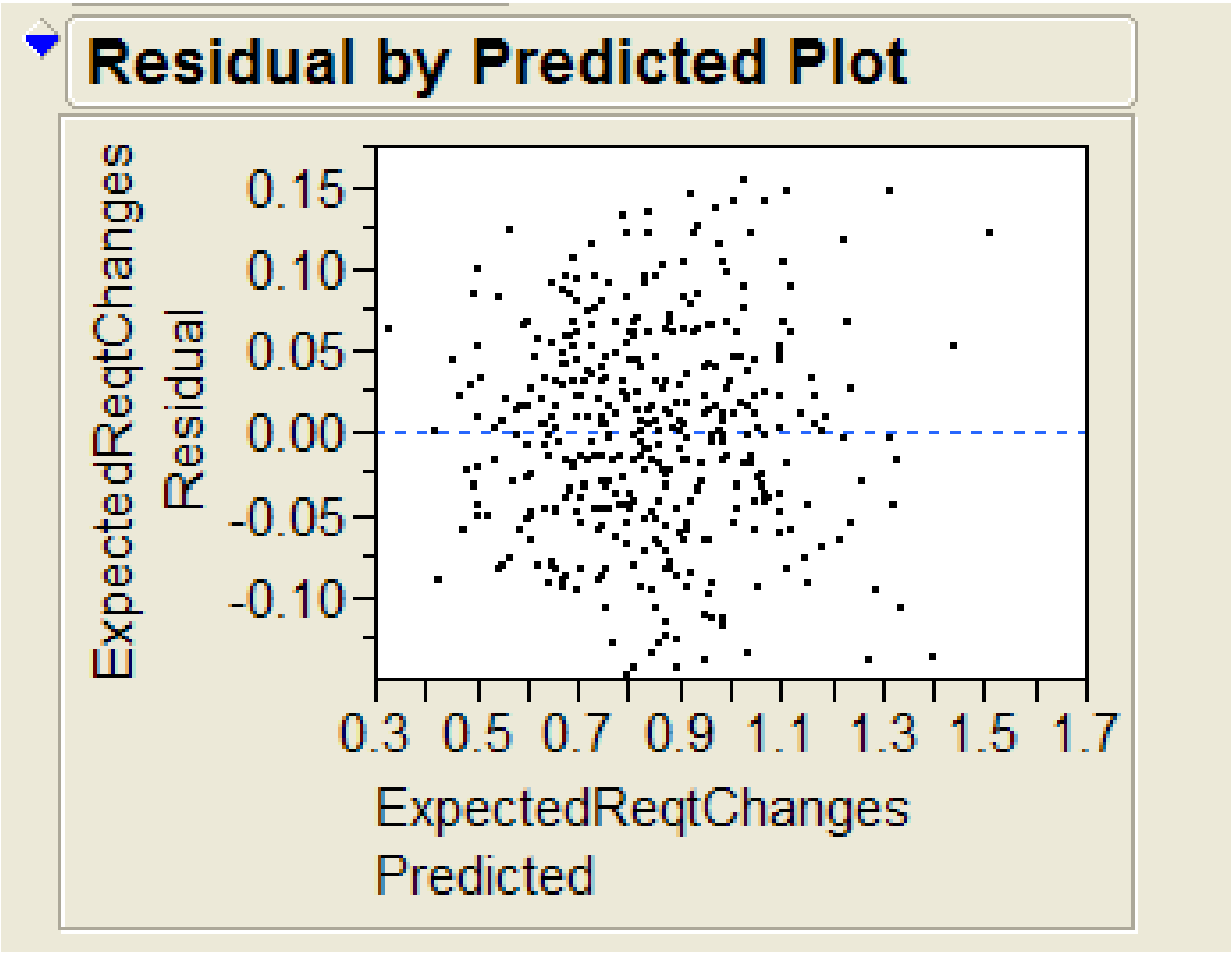
| | |
|---|---|
| RSquare | 0.903611 |
| RSquare Adj | 0.902881 |
| Root Mean Square Error | 0.063191 |
| Mean of Response | 0.844075 |
| Observations (or Sum Wgts) | 400 |

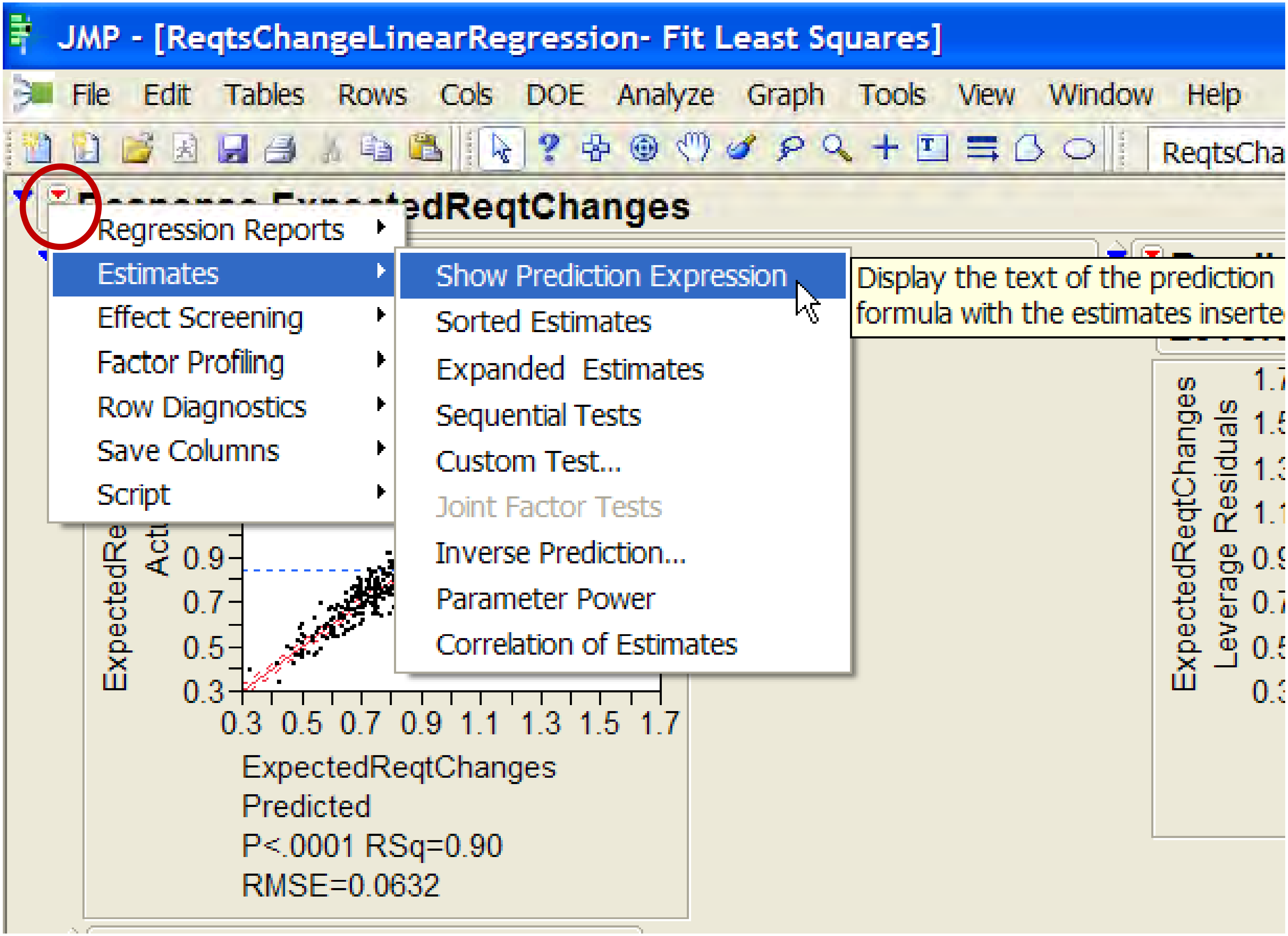
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 3 | 14.823797 | 4.94127 | 1237.457 |
| Error | 396 | 1.581261 | 0.00399 | **Prob > F** |
| C. Total | 399 | 16.405058 | | <.0001* |

## Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 0.079986 | 0.063656 | 1.26 | 0.2097 |
| ReadingLevel | 0.1576274 | 0.002628 | 59.99 | <.0001* |
| ReqtAnalystExperience | -0.003749 | 0.000859 | -4.36 | <.0001* |
| TimeSinceCustStaffChange | -0.058869 | 0.005745 | -10.25 | <.0001* |


Residual by Predicted Plot
ExpectedReqtChanges
Residual
0.15
0.10
0.05
0.00
-0.05
-0.10
0.3
0.5
0.7
0.9
1.1
1.3
1.5
1.7
ExpectedReqtChanges
Predicted


JMP - [ReqtsChangeLinearRegression- Fit Least Squares]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
ReqtsCha
edReqtChanges
Regression Reports
Estimates
Effect Screening
Factor Profiling
Row Diagnostics
Save Columns
Script
Show Prediction Expression
Sorted Estimates
Expanded Estimates
Sequential Tests
Custom Test...
Joint Factor Tests
Inverse Prediction...
Parameter Power
Correlation of Estimates
Display the text of the prediction
formula with the estimates inserte
0.3 0.5 0.7 0.9 1.1 1.3 1.5 1.7
ExpectedReqtChanges
Predicted
P<.0001 RSq=0.90
RMSE=0.0632
ExpectedReqtChanges
Leverage Residuals

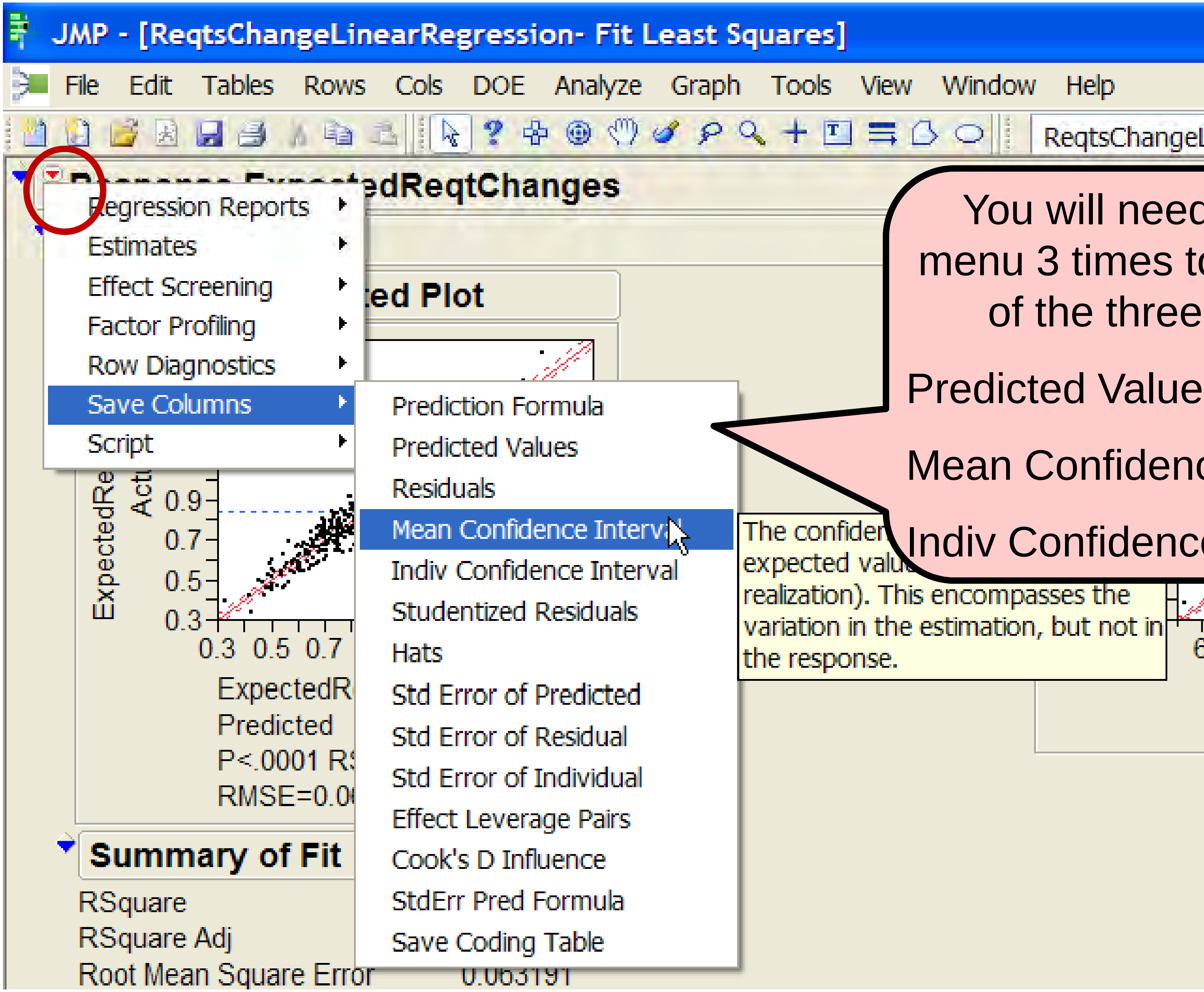
## Prediction Expression

0.0799860016248

+0.15762743466752*ReadingLevel

+ -0.0037486144177 *ReqtAnalystExperience

+ -0.0588691875889 *TimeSinceCustStaffChange




You will need to do this menu 3 times to make each of the three choices:

Predicted Values

Mean Confidence Intervals

Indiv Confidence Intervals

| | Predicted ExpectedReqtChan... | Lower 95% Mean ExpectedReqtChanges | Upper 95% Mean ExpectedReqtChanges | Lower 95% Indiv ExpectedReqtChanges | Upper 95% Indiv ExpectedReqtChanges |
|---|---|---|---|---|---|
| 1 | 0.55494838 | 0.54354212 | 0.56635463 | 0.43019439 | 0.67970237 |
| 2 | 0.50119264 | 0.48071519 | 0.52167008 | 0.37528481 | 0.62710046 |
| 3 | 0.78773032 | 0.77996762 | 0.79549302 | 0.66325657 | 0.91220407 |
| 4 | 0.65375276 | 0.64466949 | 0.66283604 | 0.52918968 | 0.77831584 |
| 5 | 0.93427302 | 0.92589222 | 0.94265383 | 0.80975919 | 1.05878685 |
| 6 | 0.83502362 | 0.82398931 | 0.84605794 | 0.71030309 | 0.95974416 |
| 7 | 0.79677671 | 0.78830532 | 0.8052481 | 0.67225675 | 0.92129667 |
| 8 | 1.0507183 | 1.0333979 | 1.06803869 | 0.92528524 | 1.17615135 |
| 9 | 0.92224158 | 0.91497575 | 0.92950742 | 0.79779783 | 1.04668534 |
| 10 | 0.49952824 | 0.48310202 | 0.51595446 | 0.37421553 | 0.62484096 |
| 11 | 0.75985048 | 0.74183342 | 0.77786753 | 0.63431933 | 0.88538163 |
| 12 | 0.75034912 | 0.7414727 | 0.75922554 | 0.62580095 | 0.87489729 |
| 13 | 0.75073824 | 0.7338668 | 0.76760968 | 0.62536639 | 0.87611009 |
| 14 | 0.99346413 | 0.98080803 | 1.00612023 | 0.86858966 | 1.1183386 |
| 15 | 0.65996448 | 0.64952963 | 0.67039932 | 0.53529555 | 0.78463341 |
| 16 | 1.27014006 | 1.25421652 | 1.28606359 | 1.14489224 | 1.39538787 |
| 17 | 0.83170791 | 0.82004771 | 0.84336812 | 0.70693045 | 0.95648538 |
| 18 | 0.86004345 | 0.84999154 | 0.87009535 | 0.73540599 | 0.98468091 |
| 19 | 0.87250377 | 0.85688522 | 0.88812231 | 0.74729436 | 0.99771317 |
| 20 | 0.65187676 | 0.6392197 | 0.66453382 | 0.5270022 | 0.77675133 |
| 21 | 0.70233417 | 0.69025672 | 0.71441163 | 0.57751703 | 0.82715132 |
| 22 | 0.81102914 | 0.7987054 | 0.82335289 | 0.68618792 | 0.93587036 |

# PPM Exercise 4: Predicting Delivered Defects with Dummy Variable Regression

# Dummy Variable Regression

The purpose of Dummy Variable Regression is to predict a continuous Y outcome using a combination of continuous and discrete x factors.

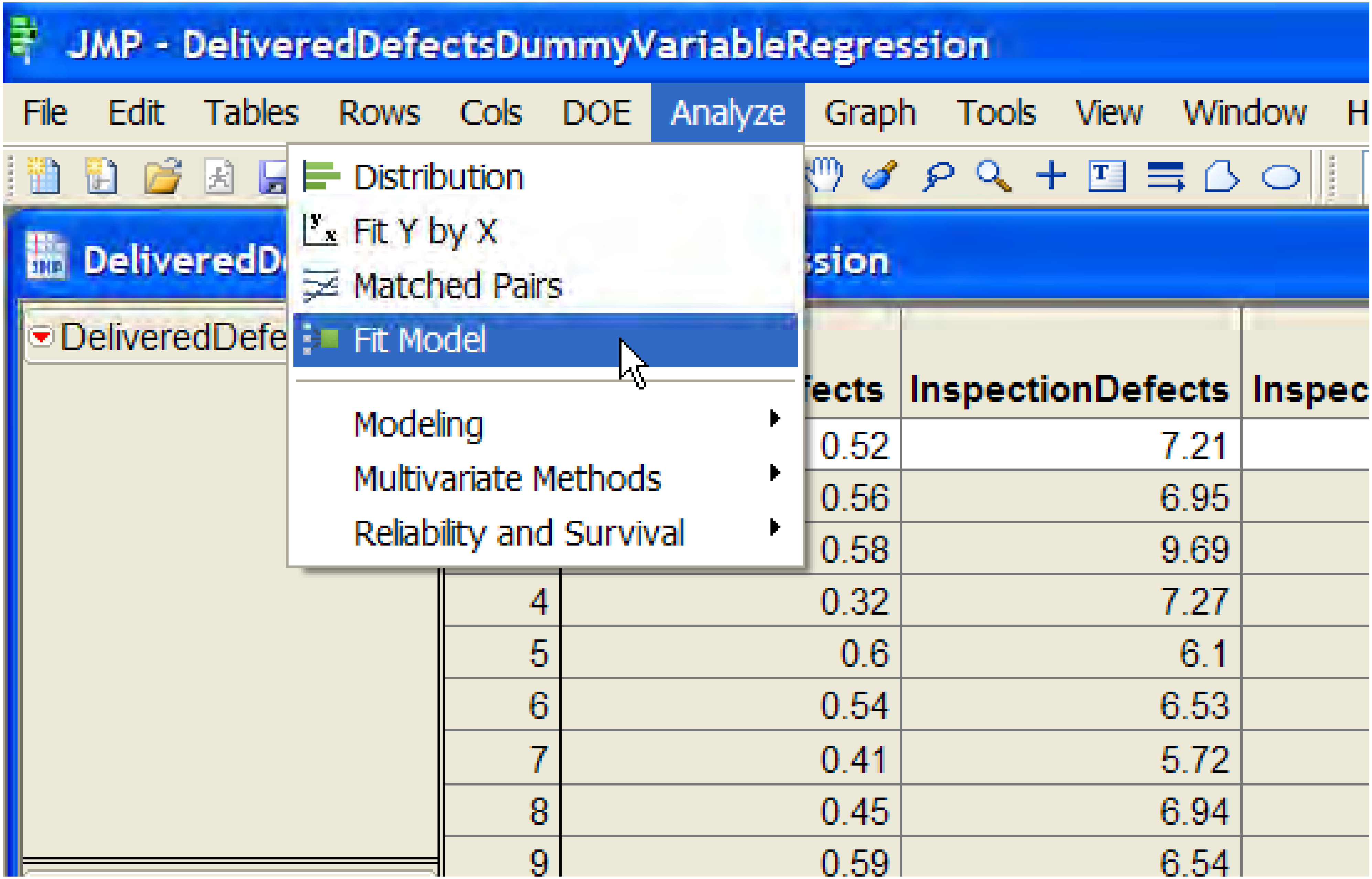
# Statistical Regression Analysis

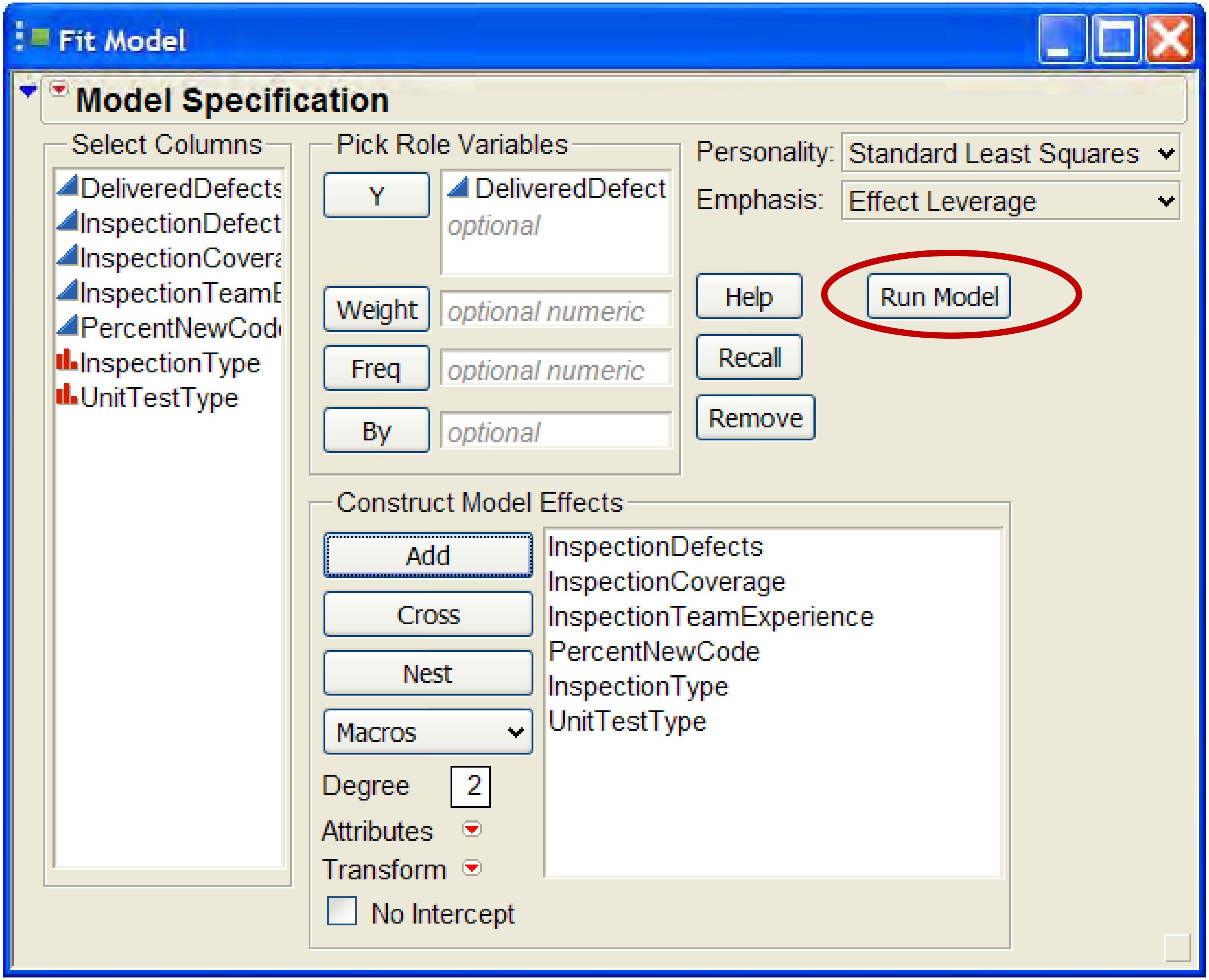
| | Y: Continuous | Y: Discrete |
|---|---|---|
| X: Discrete | ANOVA and Dummy Variable Regression | Chi-Square, Logit & Logistic Regression |
| X: Continuous | Correlation & Linear Regression | Logistic Regression |

# Open the DeliveredDefectsDummyVariableRegression.jmp file

| DeliveredDefects | InspectionDefects | InspectionCoverage | InspectionTeamExperience | PercentNewCode | InspectionType | UnitTestType |
|---|---|---|---|---|---|---|
| 0.52 | 7.21 | 81.46 | 52.55 | 75.18 | 0 | 0 |
| 0.56 | 6.95 | 81.77 | 46.76 | 76.68 | 0 | 0 |
| 0.58 | 9.69 | 81.94 | 48.12 | 77.43 | 0 | 0 |
| 0.32 | 7.27 | 90.75 | 36.93 | 75.45 | 1 | 0 |
| 0.6 | 6.1 | 74.11 | 36.5 | 77.59 | 0 | 0 |
| 0.54 | 6.53 | 76.42 | 45.37 | 78 | 0 | 0 |
| 0.41 | 5.72 | 79.52 | 31.42 | 78.35 | 0 | 1 |
| 0.45 | 6.94 | 88.68 | 48.64 | 81.13 | 0 | 0 |
| 0.59 | 6.54 | 70.69 | 27.42 | 76.8 | 0 | 0 |
| 0.6 | 6.3 | 78.14 | 22.48 | 75.93 | 0 | 0 |
| 0.55 | 6.52 | 76.01 | 33.9 | 82.39 | 0 | 0 |
| 0.6 | 9.67 | 77 | 49.69 | 80.04 | 0 | 0 |
| 0.66 | 6.54 | 65.28 | 45.86 | 76.84 | 1 | 0 |
| 0.6 | 8.45 | 79.48 | 40.3 | 80.81 | 0 | 0 |
| 0.53 | 5.77 | 77.15 | 38.23 | 78.51 | 0 | 0 |
| 0.52 | 5.77 | 78.98 | 39.47 | 79.84 | 0 | 0 |
| 0.55 | 5.68 | 73.57 | 29.03 | 80.74 | 0 | 0 |
| 0.41 | 8.57 | 80.08 | 24.82 | 76.55 | 0 | 1 |
| 0.45 | 8.99 | 85.46 | 34.95 | 80.09 | 0 | 1 |
| 0.55 | 6.38 | 76.66 | 27.05 | 82.46 | 0 | 0 |
| 0.43 | 5.83 | 80.14 | 33.77 | 78.74 | 1 | 1 |

| Factor | Role | Data Type | Description |
|---|---|---|---|
| DeliveredDefects | Y Outcome | Continuous | Delivered Defect Density normalized to KSLOC for a given feature |
| InspectionDefects | X1 Factor | Continuous | Inspection Defect Density normalized to KSLOC for a given feature |
| InspectionCoverage | X2 Factor | Continuous | The percentage of inspection criteria implemented across the code files for a given feature |
| InspectionTeamExperience | X3 Factor | Continuous | The average domain experience in months of the participants of the peer review of the feature |
| PercentNewCode | X4 Factor | Continuous | The percent of new code within the feature |
| InspectionType | X5 Factor | Nominal | A factor which reflects whether an informal peer review (0) vs a formal inspection (1) occurred for the feature |
| UnitTestType | X6 Factor | Nominal | A factor which reflects whether informal unit testing (0) vs formal unit testing (1) occurred for the feature |


JMP - DeliveredDefectsDummyVariableRegression
File
Edit
Tables
Rows
Cols
DOE
Analyze
Graph
Tools
View
Window
Distribution
Fit Y by X
Matched Pairs
Fit Model
Modeling
Multivariate Methods
Reliability and Survival
InspectionDefects
0.52 7.21
0.56 6.95
0.58 9.69
4 0.32 7.27
5 0.6 6.1
6 0.54 6.53
7 0.41 5.72
8 0.45 6.94
9 0.59 6.54



Fit Model
Model Specification
Select Columns
DeliveredDefects
InspectionDefect
InspectionCovera
InspectionTeamE
PercentNewCod
InspectionType
UnitTestType
Pick Role Variables
Y
DeliveredDefect
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
InspectionDefects
InspectionCoverage
InspectionTeamExperience
PercentNewCode
InspectionType
UnitTestType

## Summary of Fit

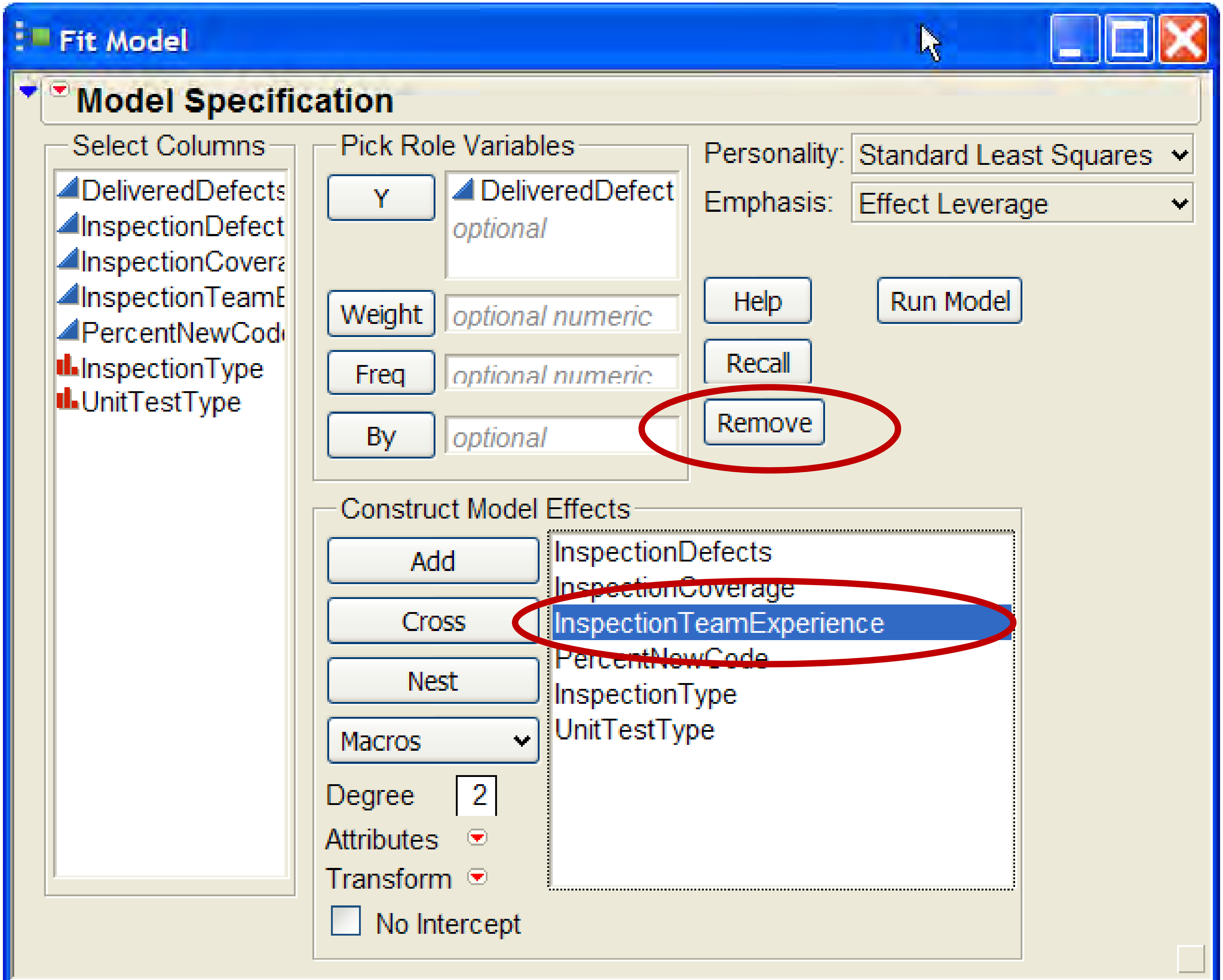
| | |
|---|---|
| RSquare | 0.846337 |
| RSquare Adj | 0.844753 |
| Root Mean Square Error | 0.033605 |
| Mean of Response | 0.483701 |
| Observations (or Sum Wgts) | 589 |

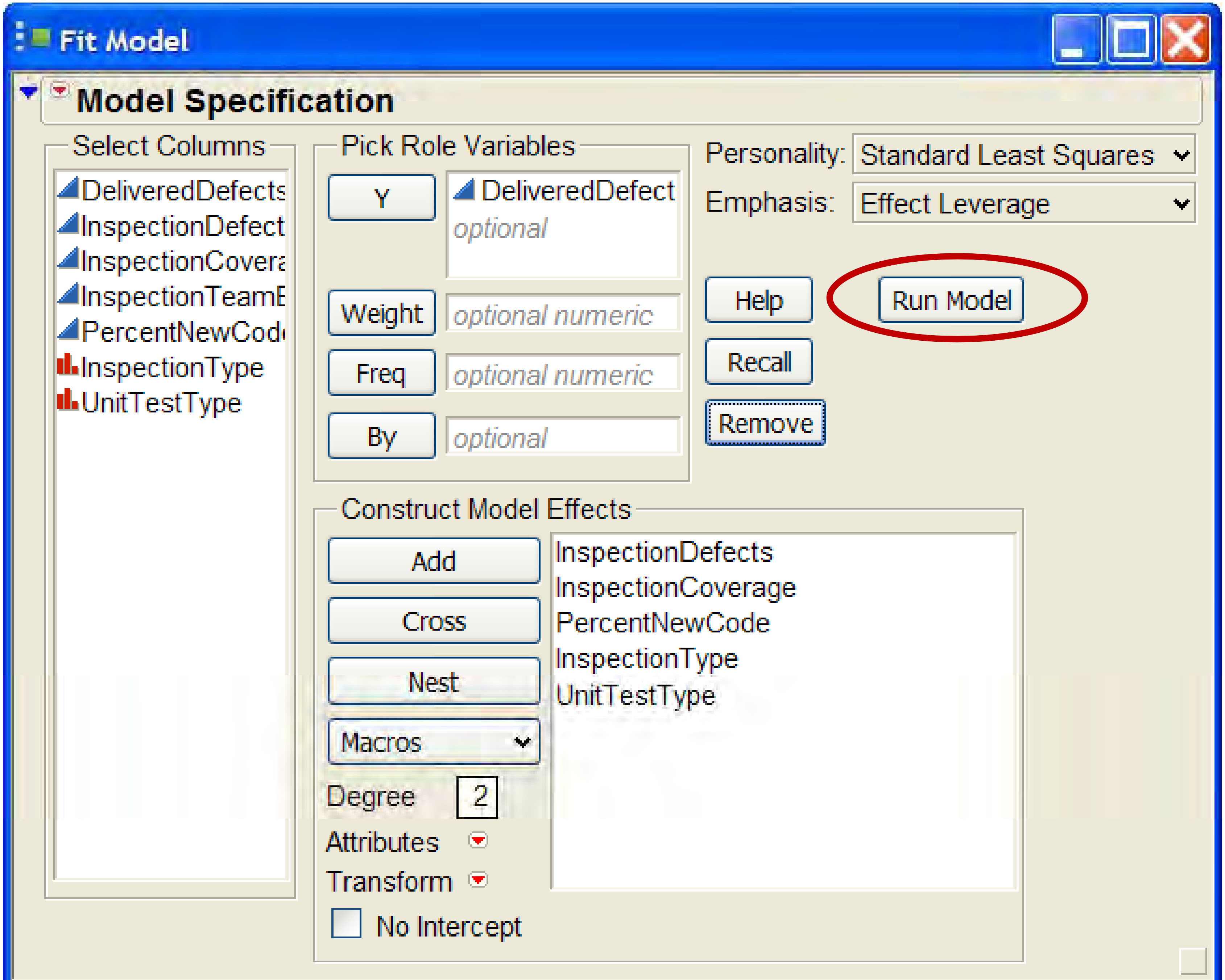
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 6 | 3.6198962 | 0.603316 | 534.2531 |
| Error | 582 | 0.6572352 | 0.001129 | **Prob > F** |
| C. Total | 588 | 4.2771314 | | <.0001* |

## Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 1.0617488 | 0.061206 | 17.35 | <.0001* |
| InspectionDefects | 0.0098332 | 0.001301 | 7.56 | <.0001* |
| InspectionCoverage | -0.009272 | 0.000278 | -33.39 | <.0001* |
| InspectionTeamExperience | -0.000149 | 0.000132 | -1.13 | 0.2588 |
| PercentNewCode | 0.0009154 | 0.000705 | 1.30 | 0.1947 |
| InspectionType[0] | 0.023847 | 0.001518 | 15.71 | <.0001* |
| UnitTestType[0] | 0.0609747 | 0.00147 | 41.47 | <.0001* |


Fit Model
Model Specification
Select Columns
DeliveredDefects
InspectionDefect
InspectionCovera
InspectionTeamE
PercentNewCod
InspectionType
UnitTestType
Pick Role Variables
Y
DeliveredDefect
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality: Standard Least Squares
Emphasis: Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree 2
Attributes
Transform
No Intercept
InspectionDefects
InspectionCoverage
InspectionTeamExperience
PercentNewCode
InspectionType
UnitTestType



Fit Model
Model Specification
Select Columns
DeliveredDefects
InspectionDefect
InspectionCovera
InspectionTeamE
PercentNewCod
InspectionType
UnitTestType
Pick Role Variables
Y
DeliveredDefect
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
InspectionDefects
InspectionCoverage
PercentNewCode
InspectionType
UnitTestType

## Summary of Fit

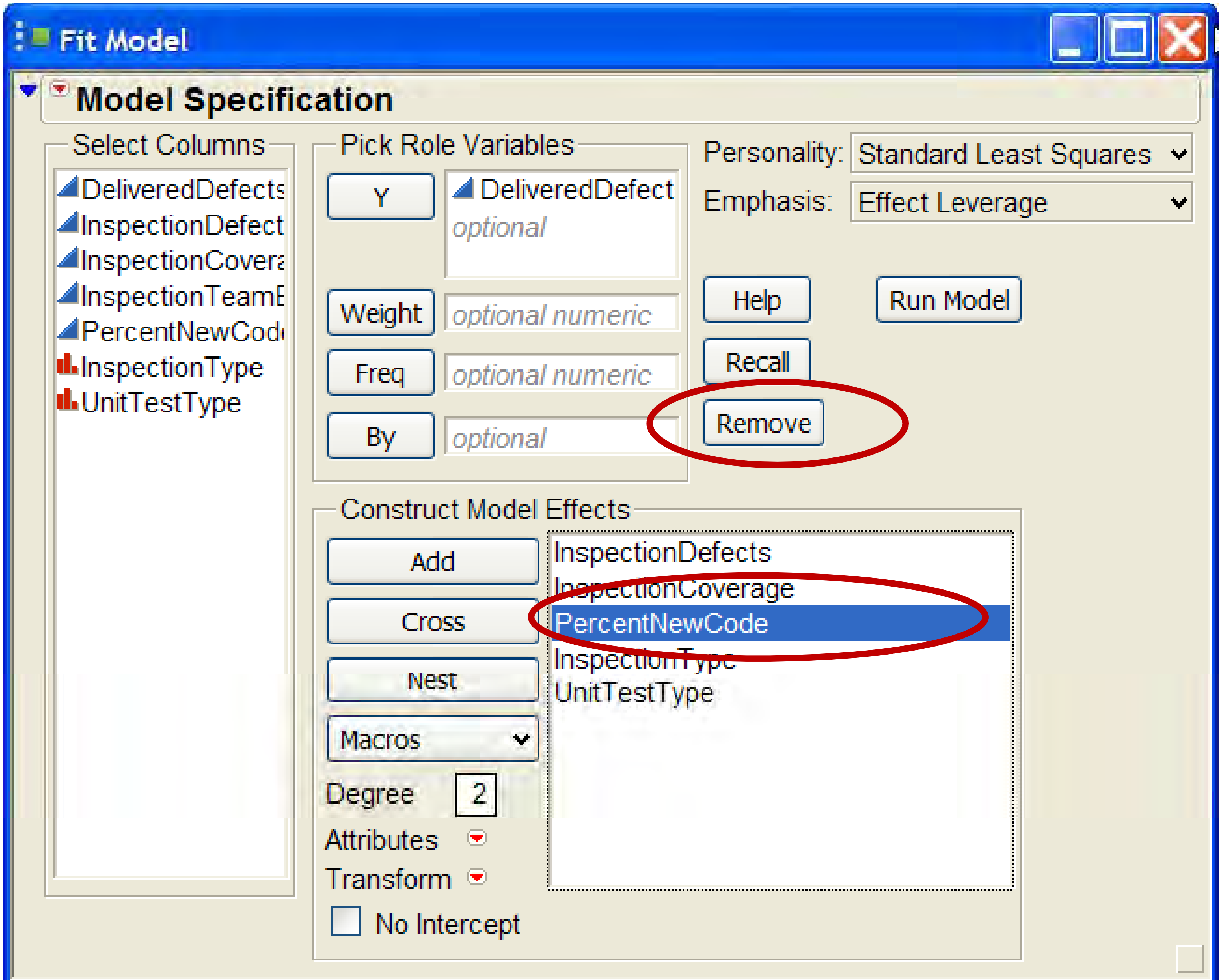
| | |
|---|---|
| RSquare | 0.846 |
| RSquare Adj | 0.844679 |
| Root Mean Square Error | 0.033613 |
| Mean of Response | 0.483701 |
| Observations (or Sum Wgts) | 589 |

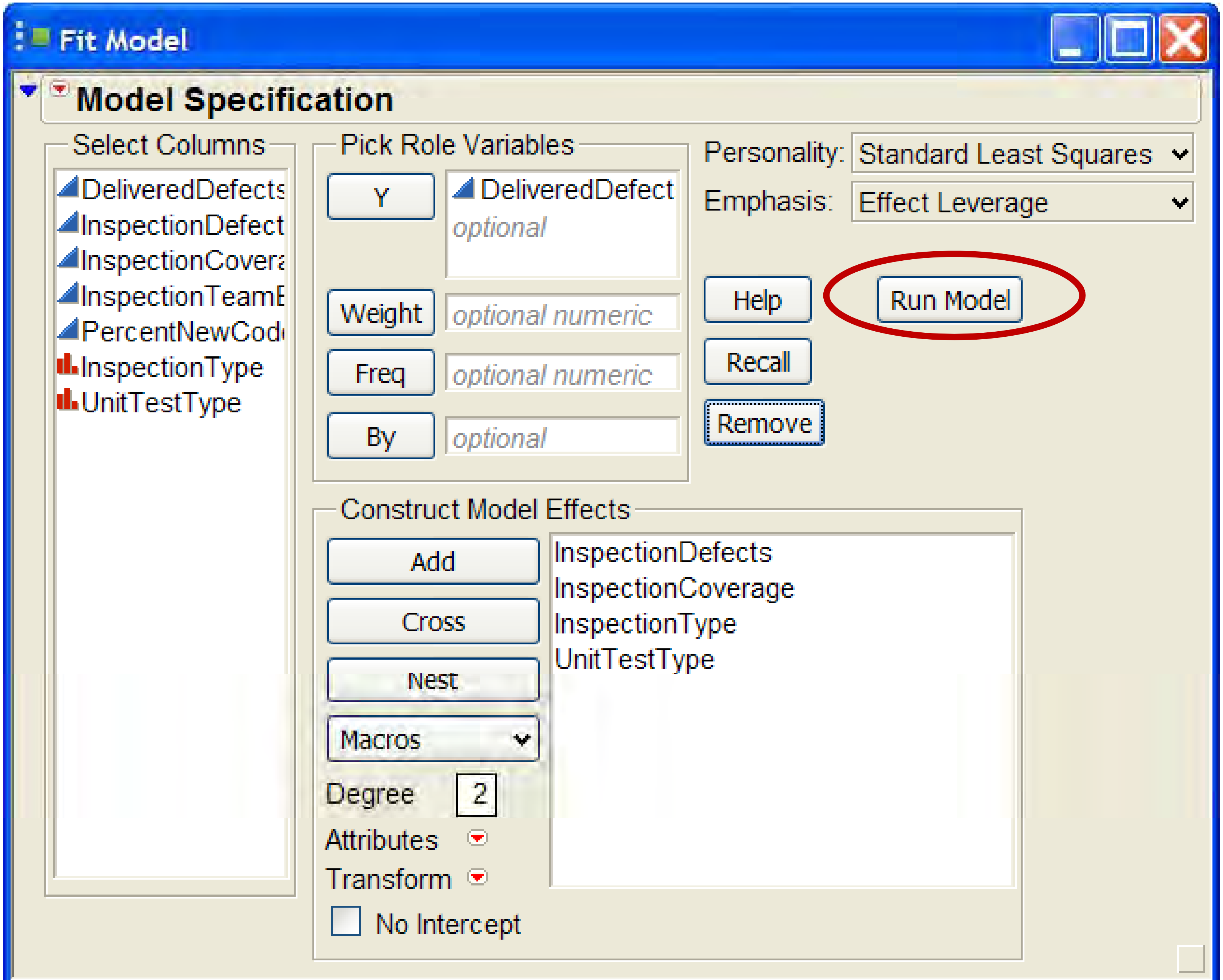
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 5 | 3.6184537 | 0.723691 | 640.5434 |
| Error | 583 | 0.6586777 | 0.001130 | **Prob > F** |
| C. Total | 588 | 4.2771314 | | <.0001* |

## Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 1.0562782 | 0.061028 | 17.31 | <.0001* |
| InspectionDefects | 0.009816 | 0.001302 | 7.54 | <.0001* |
| InspectionCoverage | -0.009266 | 0.000278 | -33.37 | <.0001* |
| PercentNewCode | 0.0009039 | 0.000705 | 1.28 | 0.2004 |
| InspectionType[0] | 0.0239018 | 0.001518 | 15.75 | <.0001* |
| UnitTestType[0] | 0.0609821 | 0.00147 | 41.47 | <.0001* |


Fit Model
Model Specification
Select Columns
DeliveredDefects
InspectionDefect
InspectionCovera
InspectionTeamE
PercentNewCod
InspectionType
UnitTestType
Pick Role Variables
Y
DeliveredDefect
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
InspectionDefects
InspectionCoverage
PercentNewCode
InspectionType
UnitTestType


Fit Model
Model Specification
Select Columns
DeliveredDefects
InspectionDefect
InspectionCovera
InspectionTeamE
PercentNewCod
InspectionType
UnitTestType
Pick Role Variables
Y
DeliveredDefect
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Standard Least Squares
Emphasis:
Effect Leverage
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
InspectionDefects
InspectionCoverage
InspectionType
UnitTestType

## Summary of Fit

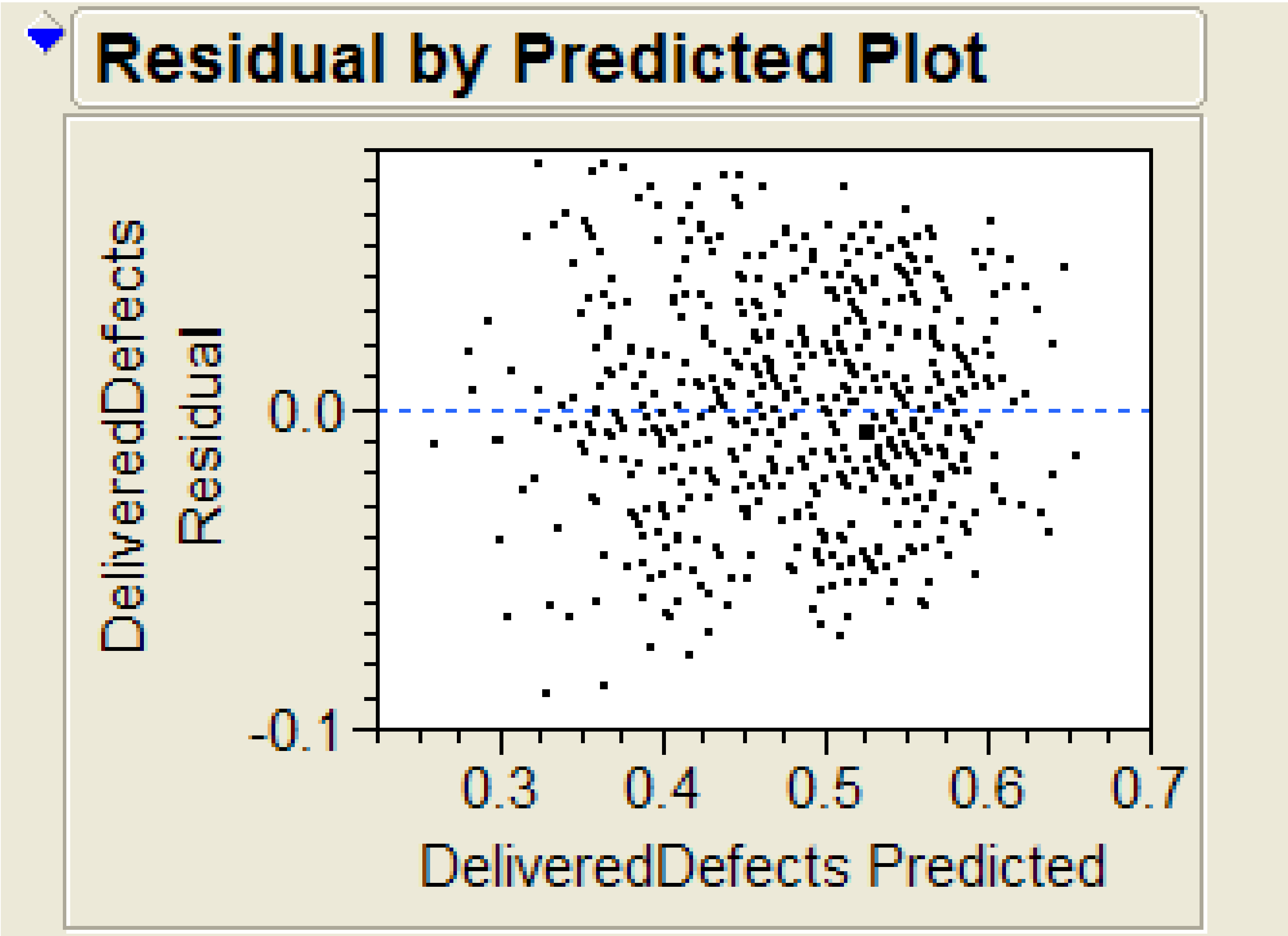
| | |
|---|---|
| RSquare | 0.845566 |
| RSquare Adj | 0.844508 |
| Root Mean Square Error | 0.033631 |
| Mean of Response | 0.483701 |
| Observations (or Sum Wgts) | 589 |

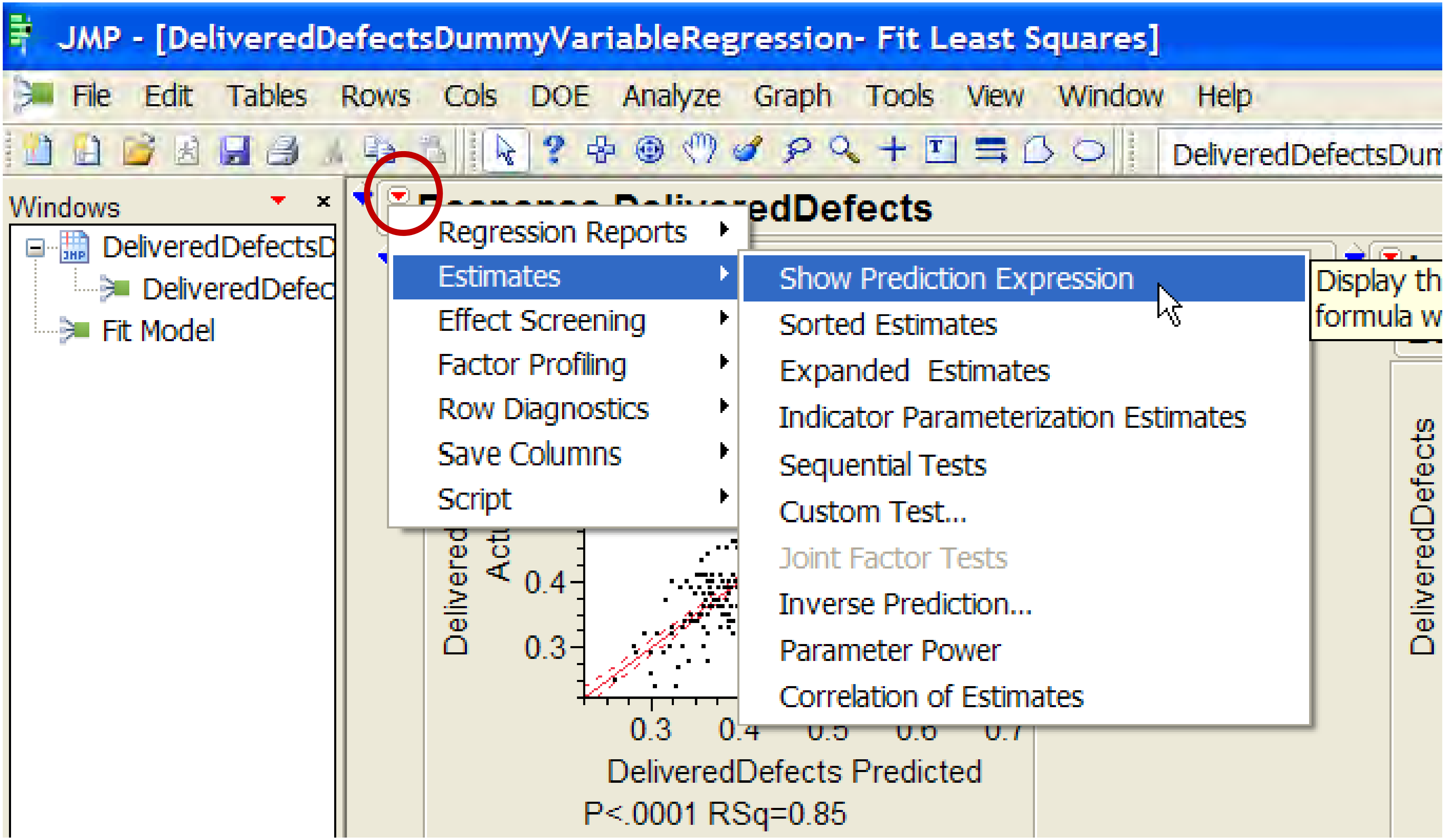
## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 4 | 3.6165970 | 0.904149 | 799.3879 |
| Error | 584 | 0.6605344 | 0.001131 | **Prob > F** |
| C. Total | 588 | 4.2771314 | | <.0001* |

## Parameter Estimates

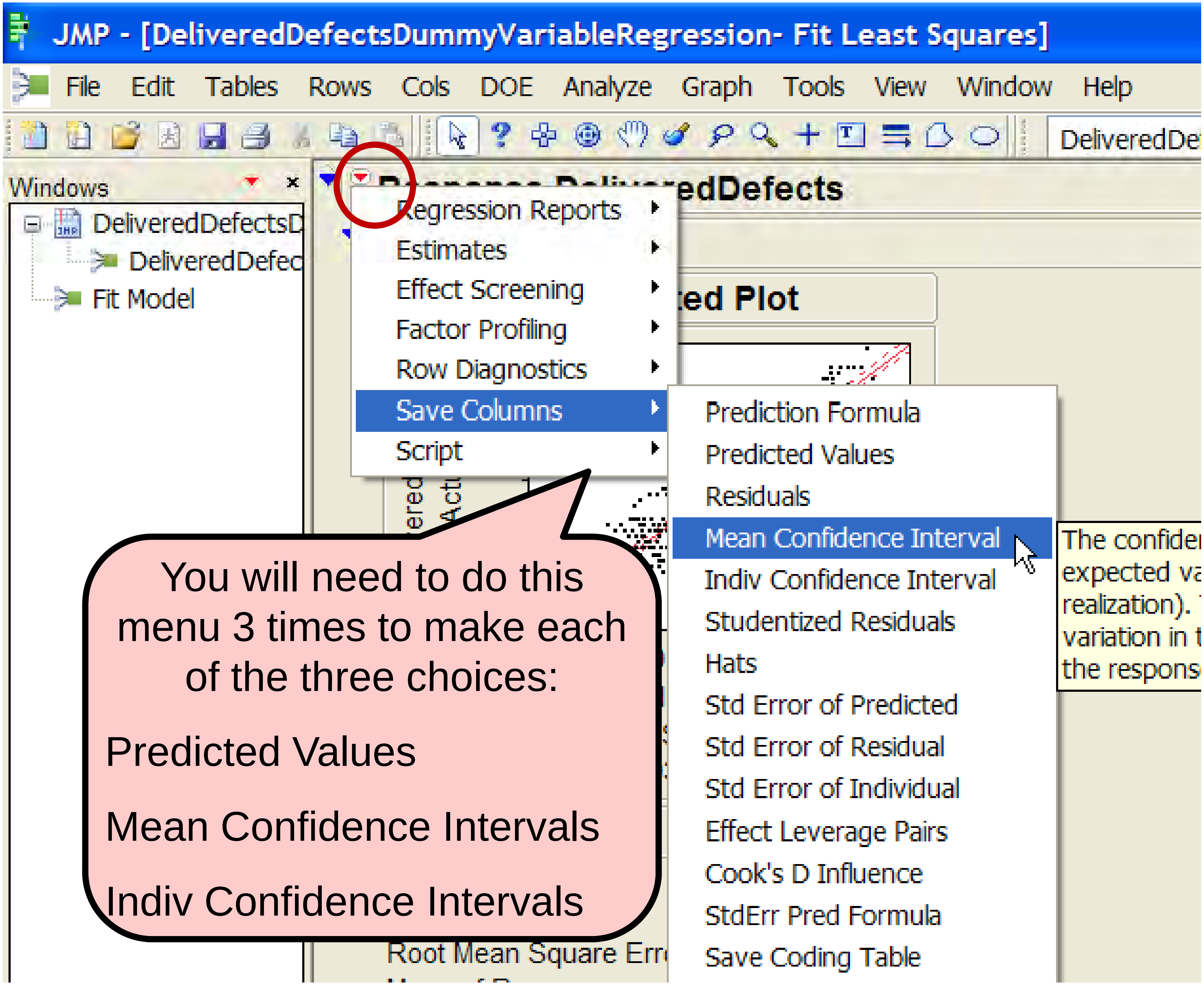
| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 1.1279627 | 0.024456 | 46.12 | <.0001* |
| InspectionDefects | 0.0097196 | 0.0013 | 7.48 | <.0001* |
| InspectionCoverage | -0.00927 | 0.000278 | -33.36 | <.0001* |
| InspectionType[0] | 0.023858 | 0.001518 | 15.71 | <.0001* |
| UnitTestType[0] | 0.0608499 | 0.001468 | 41.46 | <.0001* |


Residual by Predicted Plot
DeliveredDefects
Residual
0.0
-0.1
0.3
0.4
0.5
0.6
0.7
DeliveredDefects Predicted


JMP - [DeliveredDefectsDummyVariableRegression- Fit Least Squares]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
DeliveredDefectsDum
Windows
DeliveredDefectsD
DeliveredDefec
Fit Model
Regression Reports
Estimates
Effect Screening
Factor Profiling
Row Diagnostics
Save Columns
Script
Show Prediction Expression
Sorted Estimates
Expanded Estimates
Indicator Parameterization Estimates
Sequential Tests
Custom Test...
Joint Factor Tests
Inverse Prediction...
Parameter Power
Correlation of Estimates
Display th
formula w
DeliveredDefects Predicted
P<.0001 RSq=0.85
DeliveredDefects

## Prediction Expression

$$
\begin{aligned}
&1.12796274699737 \\
&+0.0097196166 4 * \mathrm{InspectionDefects} \\
&+\begin{array}{l}-0.0092698591556 \\ * \mathrm{InspectionCoverage}\end{array} \\
&+\mathrm{Match}\left[\mathrm{InspectionType}\right]\begin{bmatrix} 0 & \Rightarrow 0.023857998 9818 \\ 1 & \Rightarrow -0.0238579989982 \\ \mathrm{else} & \Rightarrow . \end{bmatrix} \\
&+\mathrm{Match}\left[\mathrm{UnitTestType}\right]\begin{bmatrix} 0 & \Rightarrow 0.0608498 8884934 \\ 1 & \Rightarrow -0.0608498888493 \\ \mathrm{else} & \Rightarrow . \end{bmatrix}
\end{aligned}
$$


JMP - [DeliveredDefectsDummyVariableRegression- Fit Least Squares]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
Windows
DeliveredDefectsD
DeliveredDefec
Fit Model
Regression Reports
Estimates
Effect Screening
Factor Profiling
Row Diagnostics
Save Columns
Script
Prediction Formula
Predicted Values
Residuals
Mean Confidence Interval
Indiv Confidence Interval
Studentized Residuals
Hats
Std Error of Predicted
Std Error of Residual
Std Error of Individual
Effect Leverage Pairs
Cook's D Influence
StdErr Pred Formula
Save Coding Table
You will need to do this menu 3 times to make each of the three choices:
Predicted Values
Mean Confidence Intervals
Indiv Confidence Intervals

| Predicted DeliveredDefect... | Lower 95% Mean DeliveredDefects | Upper 95% Mean DeliveredDefects | Lower 95% Indiv DeliveredDefects | Upper 95% Indiv DeliveredDefects |
|---|---|---|---|---|
| 0.52762634 | 0.52375953 | 0.53149316 | 0.46146059 | 0.5937921 |
| 0.52222559 | 0.5183626 | 0.52608858 | 0.45606005 | 0.58839112 |
| 0.54728146 | 0.53935391 | 0.55520901 | 0.48075477 | 0.61380816 |
| 0.39437653 | 0.38606592 | 0.40268715 | 0.3278031 | 0.46094996 |
| 0.58497103 | 0.57942818 | 0.59051389 | 0.51868621 | 0.65125586 |
| 0.56773709 | 0.5632922 | 0.57218199 | 0.50153504 | 0.63393915 |
| 0.40942786 | 0.40306404 | 0.41579169 | 0.34306934 | 0.47578639 |
| 0.45807366 | 0.45207219 | 0.46407514 | 0.39174891 | 0.52439842 |
| 0.62095058 | 0.61444472 | 0.62745645 | 0.55457829 | 0.68732288 |
| 0.54955743 | 0.54524504 | 0.55386981 | 0.48336413 | 0.61575072 |
| 0.57144054 | 0.56688035 | 0.57600073 | 0.50523064 | 0.63765044 |
| 0.59288017 | 0.58499115 | 0.6007692 | 0.52635806 | 0.65940229 |
| 0.62338452 | 0.6138448 | 0.63292425 | 0.55664652 | 0.69012253 |
| 0.55803299 | 0.55276691 | 0.56329907 | 0.49177073 | 0.62429525 |
| 0.55358319 | 0.54838539 | 0.55878099 | 0.48732632 | 0.61984005 |
| 0.53661935 | 0.53166763 | 0.54157106 | 0.47038133 | 0.60285736 |
| 0.58589452 | 0.57962983 | 0.59215921 | 0.51954543 | 0.65224361 |
| 0.43193765 | 0.42581413 | 0.43806118 | 0.36560174 | 0.49827356 |
| 0.38614805 | 0.37870953 | 0.39358656 | 0.31967785 | 0.45261824 |
| 0.56405439 | 0.5595407 | 0.56856808 | 0.49784768 | 0.6302611 |
| 0.35703371 | 0.35012345 | 0.36394397 | 0.29062056 | 0.42344686 |
| 0.43740385 | 0.43126051 | 0.4435472 | 0.37106611 | 0.50374159 |
| 0.39756646 | 0.39207934 | 0.40305357 | 0.33128627 | 0.46384665 |

# PPM Exercise 5: Predicting Customer Satisfaction using Ordinal Logistic Regression

# Statistical Regression Analysis

| X \ Y | Continuous | Discrete |
| --- | --- | --- |
| Discrete | ANOVA and Dummy Variable Regression | Chi-Square, Logit & Logistic Regression |
| Continuous | Correlation & Linear Regression | Logistic Regression |

# Logistic Regression

The purpose of logistic regression is to predict a discrete (attribute) Y outcome using continuous X factors.

Logistic regression belongs to the class of models generally referred to as log-linear models.

Types of logistic regression analysis include the following:

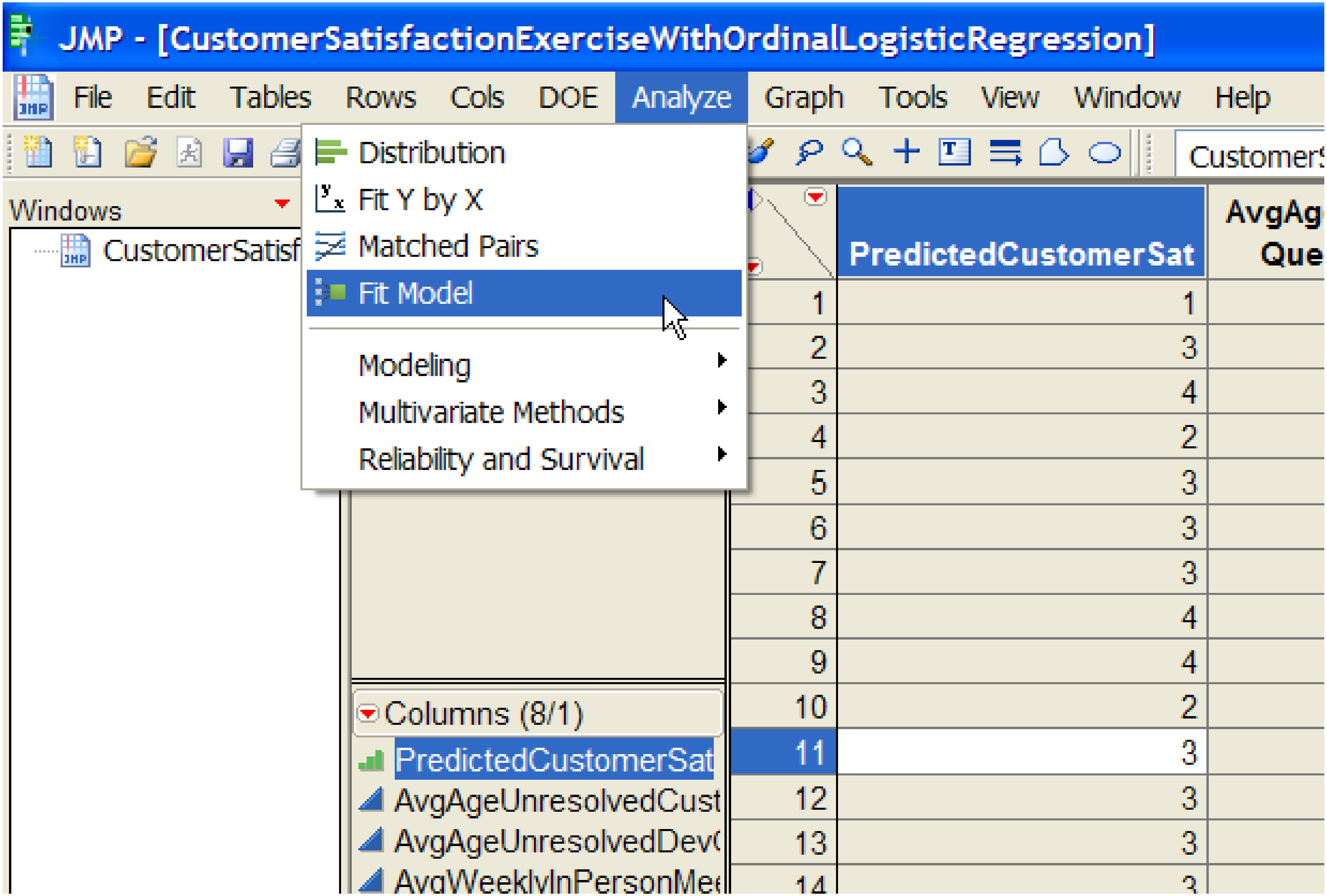
- **nominal** – a nominal Y is predicted (e.g., categorical without ordering)
- **ordinal** – an ordinal Y is predicted (e.g., categorical with ordering)
- **binary** – a binomial Y is predicted (e.g., Y is categorical with only two possible values)

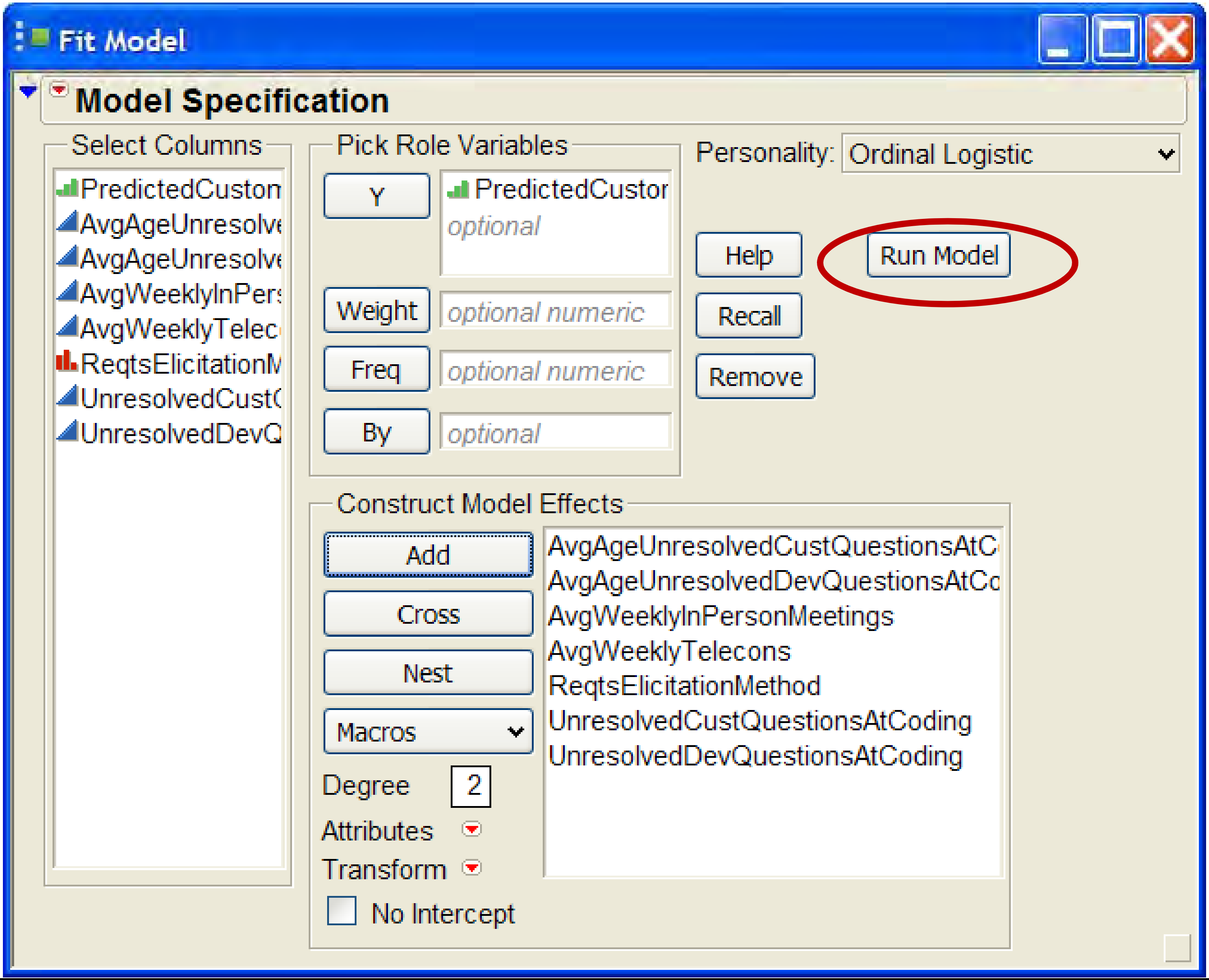
# CustomerSatisfactionExerciseWithOrdinalLogisticRegression.jmp file

| PredictedCustomerSat | AvgAgeUnresolvedCust QuestionsAtCoding | AvgAgeUnresolvedDev QuestionsAtCoding | AvgWeeklyInPerson Meetings | AvgWeekly Telecons |
|---|---|---|---|---|
| 1 | 20.27 | 49.02 | 0.15 | 4.91 |
| 3 | 19.86 | 47.53 | 0.99 | 4.15 |
| 4 | 19.52 | 51.5 | 3.57 | 3.14 |
| 2 | 18.56 | 49.02 | 0.59 | 4.58 |
| 3 | 20.45 | 46.22 | 1.34 | 2.48 |
| 3 | 20.4 | 48.22 | 2.49 | 0.84 |
| 3 | 19.42 | 45.43 | 0.41 | 4.02 |
| 4 | 19.85 | 48.03 | 1.81 | 4.69 |
| 4 | 19.98 | 47.25 | 3.13 | 3.95 |
| 2 | 19.99 | 48.93 | 1.39 | 1.67 |
| 3 | 19.11 | 47.92 | 1.52 | 1.76 |
| 3 | 20.95 | 47.93 | 0.69 | 5.17 |
| 3 | 19.5 | 41.1 | 2.8 | 5.31 |
| 3 | 19.72 | 42.94 | 1.44 | 4.18 |
| 3 | 20.69 | 44.88 | 1.27 | 2.53 |
| 2 | 18.95 | 50.49 | 0.31 | 6.12 |

| ReqtsElicitation Method | UnresolvedCustQuestions AtCoding | UnresolvedDevQuestions AtCoding |
|---|---|---|
| 1.00 | 28.8 | 36.33 |
| 2.00 | 30.88 | 41.15 |
| 1.00 | 23.74 | 47.75 |
| 1.00 | 26.16 | 49.61 |
| 2.00 | 32.35 | 43.4 |
| 1.00 | 27.56 | 42.92 |
| 3.00 | 25.53 | 45.85 |
| 2.00 | 26.22 | 47.87 |
| 1.00 | 31.26 | 42.48 |
| 1.00 | 31.62 | 41.47 |
| 1.00 | 26.34 | 46.71 |
| 2.00 | 31.38 | 51.41 |
| 1.00 | 24.47 | 45.12 |
| 1.00 | 30.6 | 44.8 |
| 2.00 | 27.59 | 47.24 |
| 1.00 | 27.6 | 48.9 |
| 1.00 | 29.18 | 42.58 |

| Factor | Role | Data Type | Description |
|---|---|---|---|
| PredictedCustomerSat | Y Outcome | Ordinal | Very Low=1; Low=2; Medium=3; High=4; Very High=5 |
| AvgAgeUnresolvedCustQuestionsAtCoding | X1 Factor | Continuous | Average Age in Work Days of Unresolved Questions From Customer at the Beginning of Coding Phase |
| AvgAgeUnresolvedDevQuestionsAtCoding | X2 Factor | Continuous | Average Age in Work Days of Unresolved Questions From Developer Team at the Beginning of Coding Phase |
| AvgWeeklyInPersonMeetings | X3 Factor | Continuous | Average Number of Face to Face meetings per week between the Development Team and the Customer |
| AvgWeeklyTelecons | X4 Factor | Continuous | Average Number of Teleconference Calls held each Week between the Development Team and the Customer |
| ReqtsElicitationMethod | X5 Factor | Nominal | Strictly Spec Driven=1; Interview=2; Prototyping=3 |
| UnResolvedCustQuestionsAtCoding | X6 Factor | Continuous | Number of Unresolved Questions From Customer at the Beginning of Coding Phase |
| UnResolvedDevQuestionsAtCoding | X7 Factor | Continuous | Number of Unresolved Questions From Developer Team at the Beginning of Coding Phase |


JMP - [CustomerSatisfactionExerciseWithOrdinalLogisticRegression]
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
Distribution
Fit Y by X
Matched Pairs
Fit Model
Modeling
Multivariate Methods
Reliability and Survival
Windows
CustomerSatisf
PredictedCustomerSat
Columns (8/1)
AvgAgeUnresolvedCust
AvgAgeUnresolvedDev
1 1
2 3
3 4
4 2
5 3
6 3
7 3
8 4
9 4
10 2
11 3
12 3
13 3


Fit Model
Model Specification
Select Columns
PredictedCuston
AvgAgeUnresolve
AvgAgeUnresolve
AvgWeeklyInPers
AvgWeeklyTelec
ReqtsElicitationM
UnresolvedCust
UnresolvedDevQ
Pick Role Variables
Y
PredictedCustor
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Ordinal Logistic
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
AvgAgeUnresolvedCustQuestionsAtC
AvgAgeUnresolvedDevQuestionsAtCo
AvgWeeklyInPersonMeetings
AvgWeeklyTelecons
ReqtsElicitationMethod
UnresolvedCustQuestionsAtCoding
UnresolvedDevQuestionsAtCoding

## Ordinal Logistic Fit for PredictedCustomerSat

### Iteration History

### Whole Model Test

| Model | -LogLikelihood | DF | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Difference | 100.56109 | 8 | 201.1222 | <.0001* |
| Full | 8.84501 | | | |
| Reduced | 109.40610 | | | |

| | |
|---|---|
| RSquare (U) | 0.9192 |
| Observations (or Sum Wgts) | 90 |

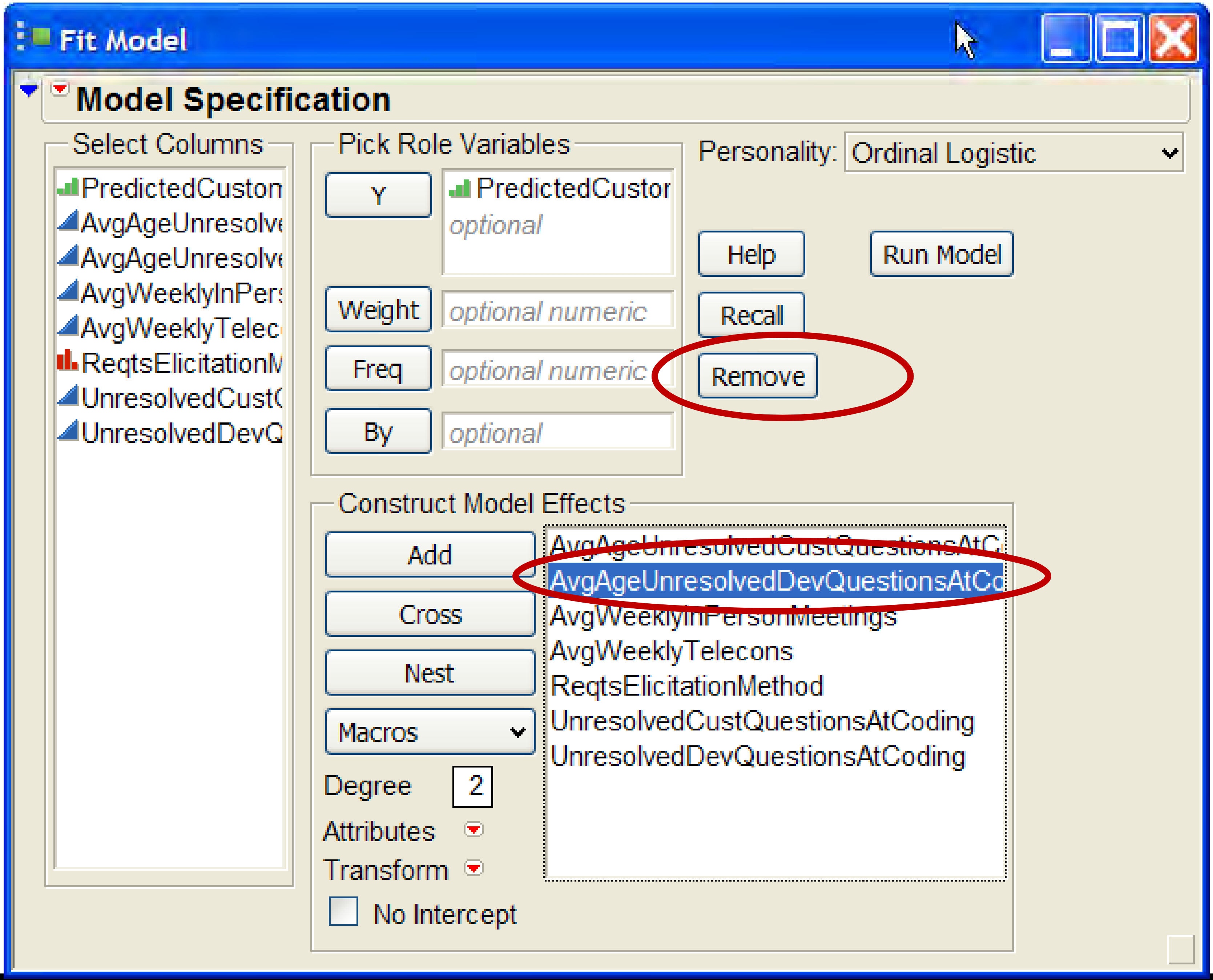
Converged by Gradient

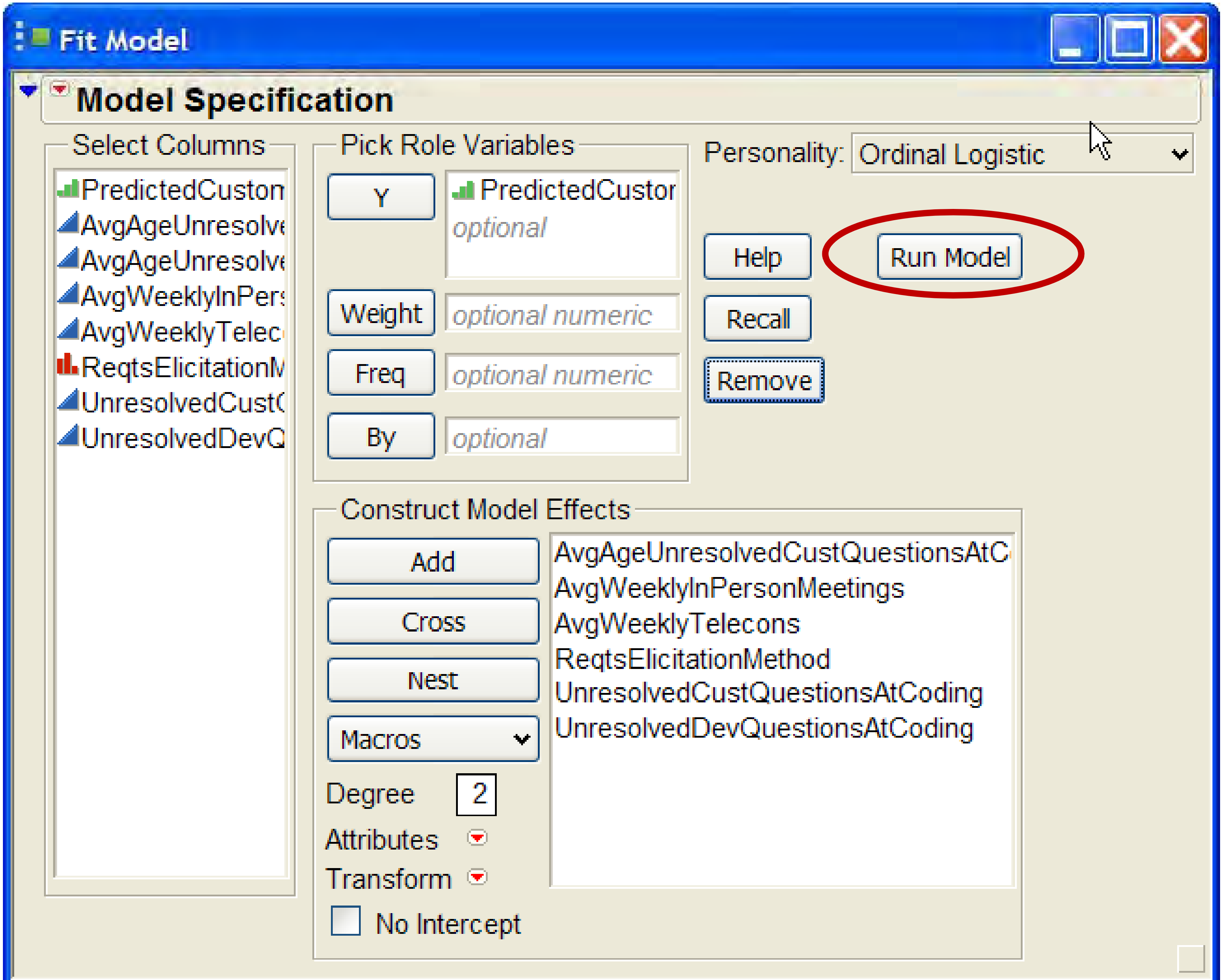
### Lack Of Fit

| Source | DF | -LogLikelihood | ChiSquare |
|---|---|---|---|
| Lack Of Fit | 348 | 8.8450140 | 17.69003 |
| Saturated | 356 | 0.0000000 | **Prob>ChiSq** |
| Fitted | 8 | 8.8450140 | 1.0000 |

## Parameter Estimates

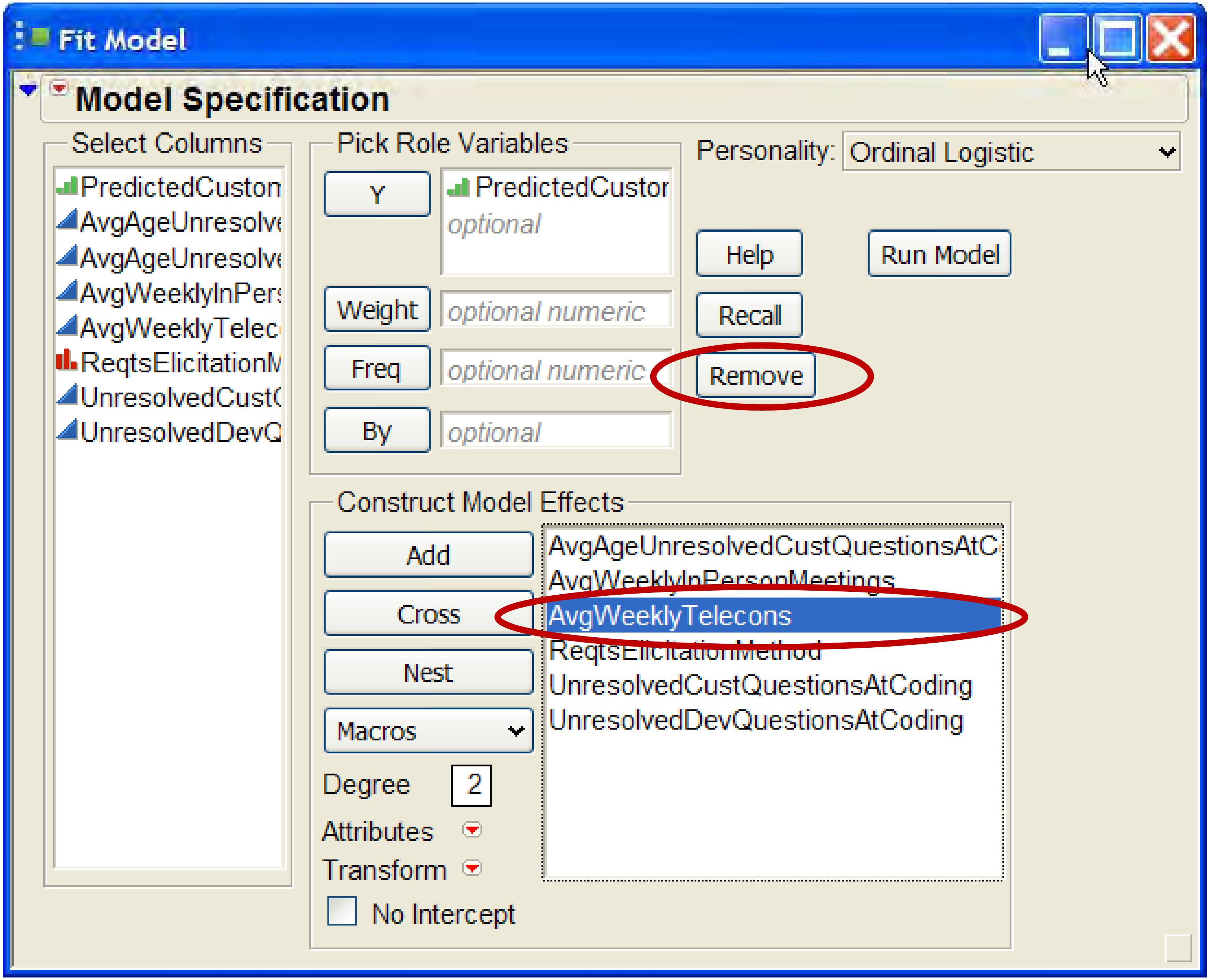
| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept[1] | 92.9334278 | 51.810259 | 3.22 | 0.0729 |
| Intercept[2] | 148.597425 | 71.058627 | 4.37 | 0.0365* |
| Intercept[3] | 203.99553 | 92.369135 | 4.88 | 0.0272* |
| Intercept[4] | 254.214635 | 112.83919 | 5.08 | 0.0243* |
| AvgAgeUnresolvedCustQuestionsAtCoding | 1.96123574 | 1.0437171 | 3.53 | 0.0602 |
| AvgAgeUnresolvedDevQuestionsAtCoding | -0.4579541 | 0.5633572 | 0.66 | 0.4163 |
| AvgWeeklyInPersonMeetings | -35.583689 | 15.138283 | 5.53 | 0.0187* |
| AvgWeeklyTelecons | -0.7033199 | 0.8018375 | 0.77 | 0.3804 |
| ReqtsElicitationMethod[1.00] | 31.0735426 | 13.061873 | 5.66 | 0.0174* |
| ReqtsElicitationMethod[2.00] | -4.398162 | 2.3425233 | 3.53 | 0.0604 |
| UnresolvedCustQuestionsAtCoding | -1.7209171 | 0.7953154 | 4.68 | 0.0305* |
| UnresolvedDevQuestionsAtCoding | -2.152774 | 0.8611309 | 6.25 | 0.0124* |

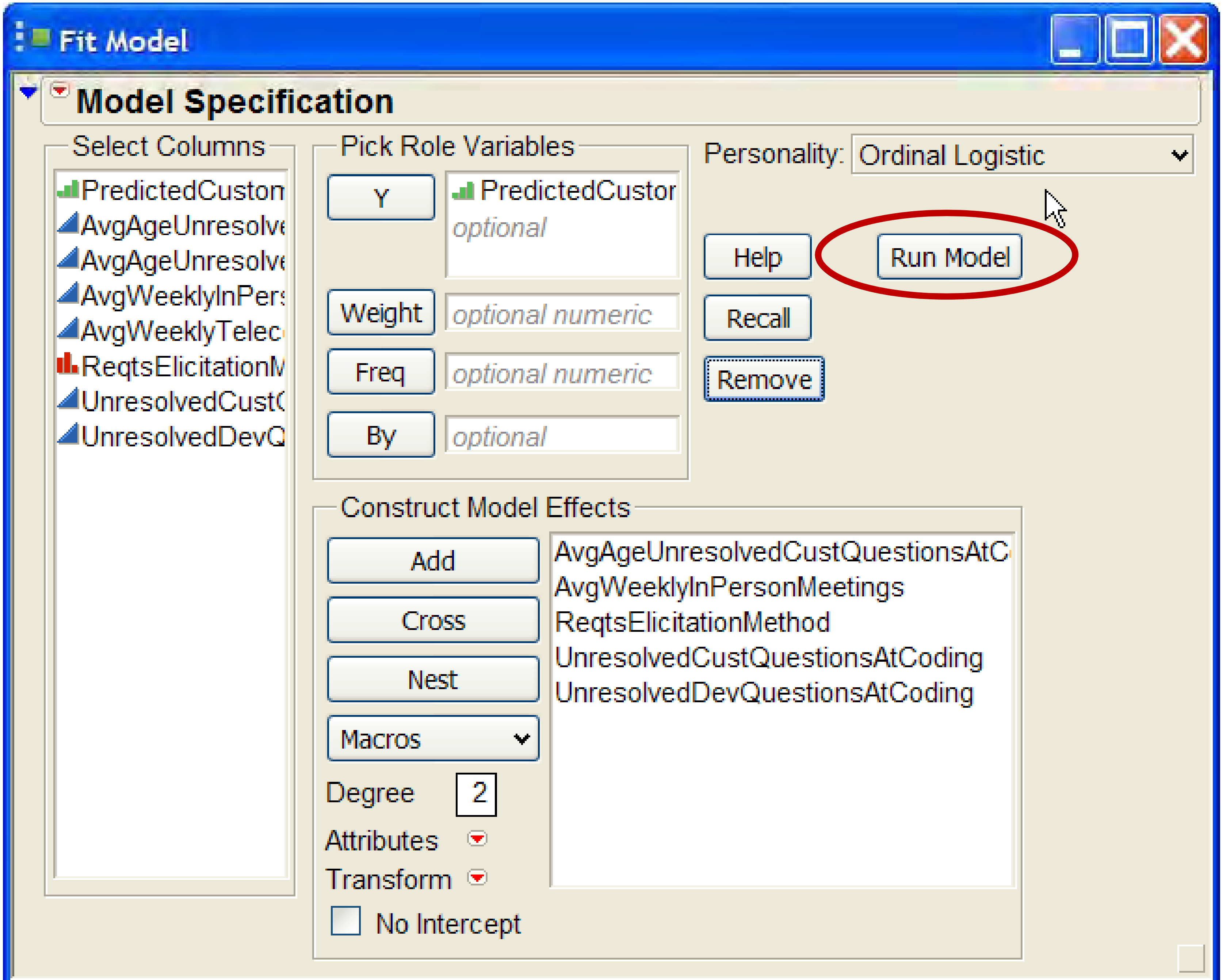

Fit Model
Model Specification
Select Columns
PredictedCuston
AvgAgeUnresolve
AvgAgeUnresolve
AvgWeeklyInPers
AvgWeeklyTelec
ReqtsElicitationM
UnresolvedCust(
UnresolvedDevQ
Pick Role Variables
Y
PredictedCustor
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Ordinal Logistic
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
AvgAgeUnresolvedCustQuestionsAtC
AvgAgeUnresolvedDevQuestionsAtCo
AvgWeeklyInPersonMeetings
AvgWeeklyTelecons
ReqtsElicitationMethod
UnresolvedCustQuestionsAtCoding
UnresolvedDevQuestionsAtCoding


Fit Model
Model Specification
Select Columns
PredictedCuston
AvgAgeUnresolve
AvgAgeUnresolve
AvgWeeklyInPers
AvgWeeklyTelec
ReqtsElicitationM
UnresolvedCust(
UnresolvedDevQ
Pick Role Variables
Y
PredictedCustor
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Ordinal Logistic
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
AvgAgeUnresolvedCustQuestionsAtC
AvgWeeklyInPersonMeetings
AvgWeeklyTelecons
ReqtsElicitationMethod
UnresolvedCustQuestionsAtCoding
UnresolvedDevQuestionsAtCoding

## Parameter Estimates

| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept[1] | 62.129936 | 27.6369 | 5.05 | 0.0246* |
| Intercept[2] | 110.845603 | 40.444952 | 7.51 | 0.0061* |
| Intercept[3] | 160.156478 | 58.481434 | 7.50 | 0.0062* |
| Intercept[4] | 203.407054 | 74.122921 | 7.53 | 0.0061* |
| AvgAgeUnresolvedCustQuestionsAtCoding | 1.66957142 | 0.8736435 | 3.65 | 0.0560 |
| AvgWeeklyInPersonMeetings | -31.445747 | 12.098374 | 6.76 | 0.0093* |
| AvgWeeklyTelecons | -0.5290472 | 0.7016369 | 0.57 | 0.4508 |
| ReqtsElicitationMethod[1.00] | 27.4484984 | 10.472716 | 6.87 | 0.0088* |
| ReqtsElicitationMethod[2.00] | -3.9627567 | 2.0628873 | 3.69 | 0.0547 |
| UnresolvedCustQuestionsAtCoding | -1.4306074 | 0.5823547 | 6.03 | 0.0140* |
| UnresolvedDevQuestionsAtCoding | -1.908446 | 0.6884446 | 7.68 | 0.0056* |


Fit Model
Model Specification
Select Columns
PredictedCuston
AvgAgeUnresolve
AvgAgeUnresolve
AvgWeeklyInPers
AvgWeeklyTelec
ReqtsElicitationM
UnresolvedCust(
UnresolvedDevQ
Pick Role Variables
Y
PredictedCustor
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Ordinal Logistic
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
AvgAgeUnresolvedCustQuestionsAtC
AvgWeeklyInPersonMeetings
AvgWeeklyTelecons
ReqtsElicitationMethod
UnresolvedCustQuestionsAtCoding
UnresolvedDevQuestionsAtCoding


Fit Model
Model Specification
Select Columns
PredictedCuston
AvgAgeUnresolve
AvgAgeUnresolve
AvgWeeklyInPer
AvgWeeklyTelec
ReqtsElicitationM
UnresolvedCust(
UnresolvedDevQ
Pick Role Variables
Y
PredictedCustor
optional
Weight
optional numeric
Freq
optional numeric
By
optional
Personality:
Ordinal Logistic
Help
Run Model
Recall
Remove
Construct Model Effects
Add
Cross
Nest
Macros
Degree
2
Attributes
Transform
No Intercept
AvgAgeUnresolvedCustQuestionsAtC
AvgWeeklyInPersonMeetings
ReqtsElicitationMethod
UnresolvedCustQuestionsAtCoding
UnresolvedDevQuestionsAtCoding

## Parameter Estimates

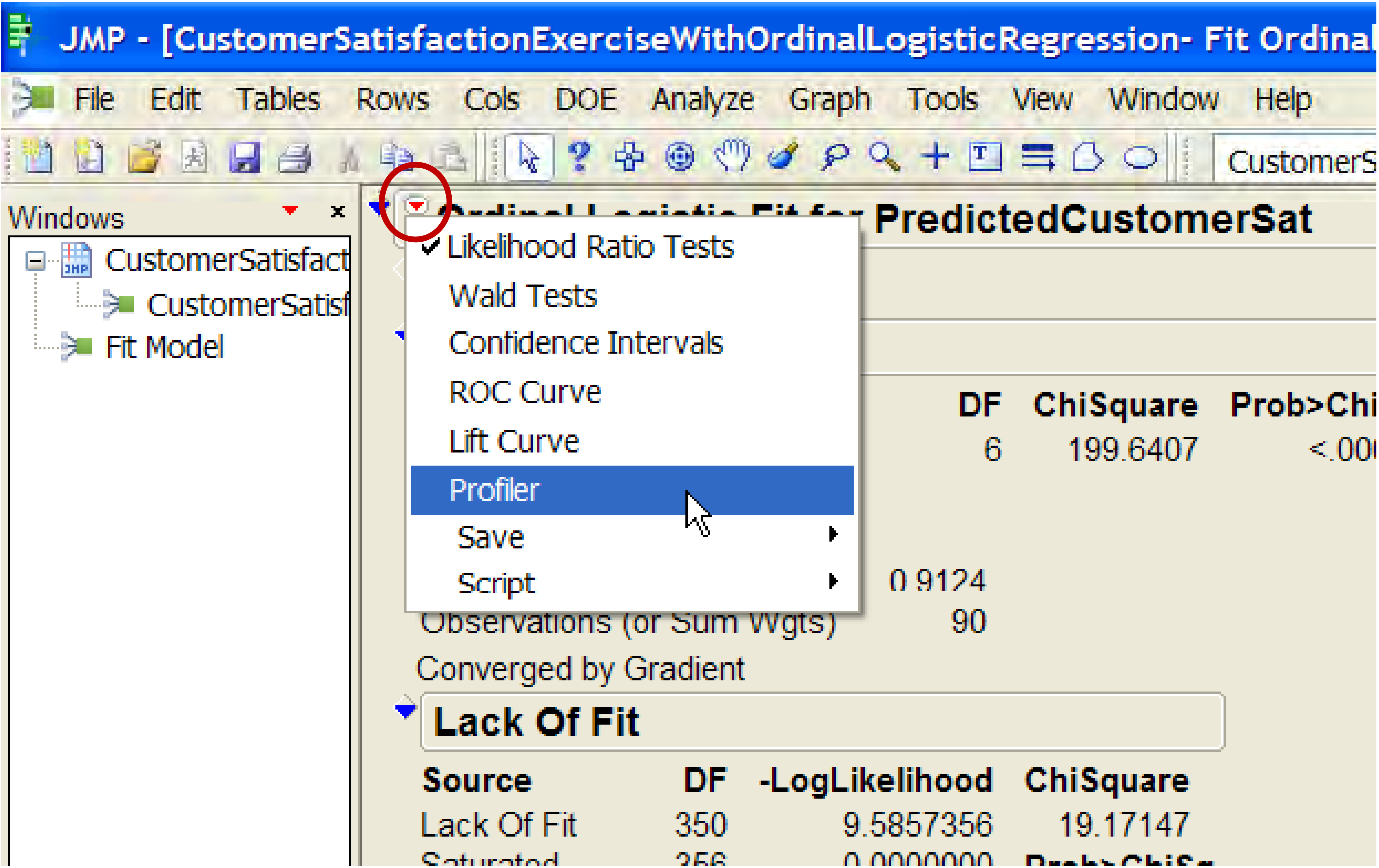
| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept[1] | 54.9797203 | 24.345705 | 5.10 | 0.0239* |
| Intercept[2] | 100.794188 | 33.940026 | 8.82 | 0.0030* |
| Intercept[3] | 145.703314 | 48.287961 | 9.10 | 0.0025* |
| Intercept[4] | 186.096476 | 61.562511 | 9.14 | 0.0025* |
| AvgAgeUnresolvedCustQuestionsAtCoding | 1.56135331 | 0.768001 | 4.13 | 0.0421* |
| AvgWeeklyInPersonMeetings | -28.947564 | 10.104355 | 8.21 | 0.0042* |
| ReqtsElicitationMethod[1.00] | 25.2121153 | 8.6077178 | 8.58 | 0.0034* |
| ReqtsElicitationMethod[2.00] | -3.5025523 | 1.7280995 | 4.11 | 0.0427* |
| UnresolvedCustQuestionsAtCoding | -1.2619774 | 0.4359331 | 8.38 | 0.0038* |
| UnresolvedDevQuestionsAtCoding | -1.8034997 | 0.6076278 | 8.81 | 0.0030* |

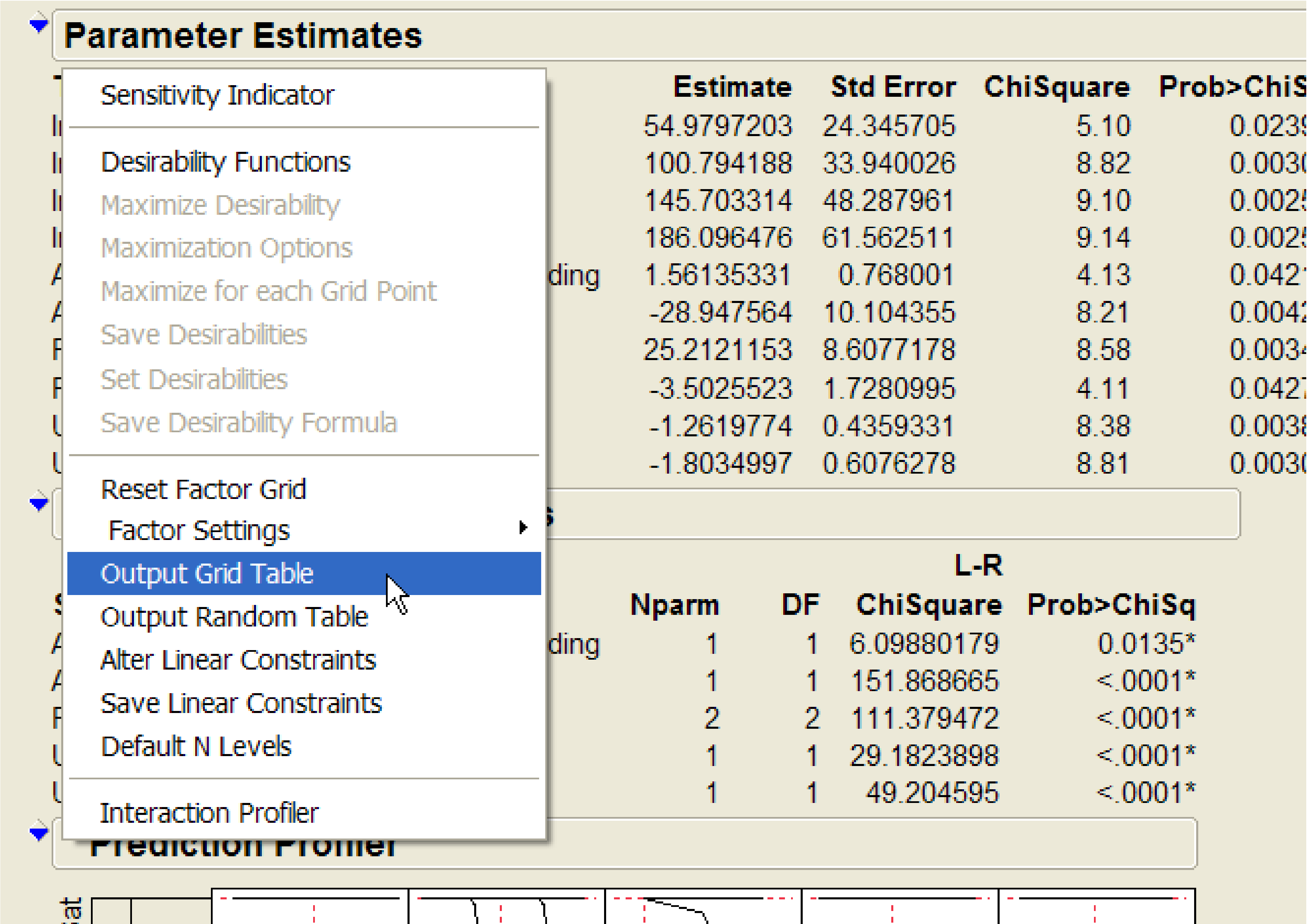
## Whole Model Test

| Model | -LogLikelihood | DF | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Difference | 99.82037 | 6 | 199.6407 | <.0001* |
| Full | 9.58574 | | | |
| Reduced | 109.40610 | | | |

| | |
|---|---|
| RSquare (U) | 0.9124 |
| Observations (or Sum Wgts) | 90 |

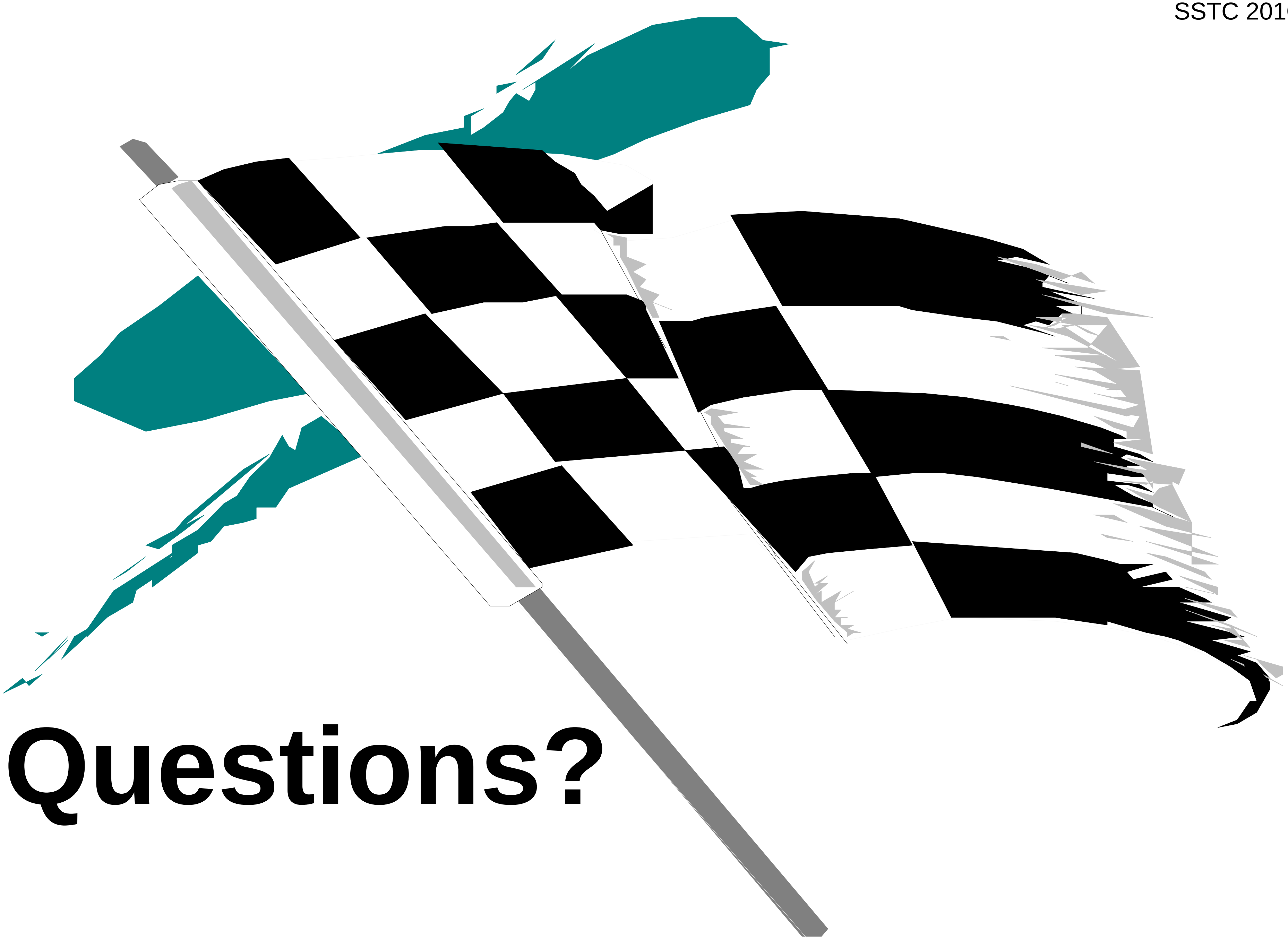
Converged by Gradient


JMP - [CustomerSatisfactionExerciseWithOrdinalLogisticRegression- Fit Ordinal
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
CustomerS
Windows
CustomerSatisfact
CustomerSatisf
Fit Model
PredictedCustomerSat
Likelihood Ratio Tests
Wald Tests
Confidence Intervals
ROC Curve
Lift Curve
Profiler
Save
Script
DF ChiSquare Prob>Chi
6 199.6407
0.9124
90
Converged by Gradient
Lack Of Fit
Source DF -LogLikelihood ChiSquare
Lack Of Fit 350 9.5857356 19.17147


Parameter Estimates
Sensitivity Indicator
Desirability Functions
Maximize Desirability
Maximization Options
Maximize for each Grid Point
Save Desirabilities
Set Desirabilities
Save Desirability Formula
Reset Factor Grid
Factor Settings
Output Grid Table
Output Random Table
Alter Linear Constraints
Save Linear Constraints
Default N Levels
Interaction Profiler
Estimate Std Error ChiSquare Prob>ChiS
54.9797203 24.345705 5.10
100.794188 33.940026 8.82
145.703314 48.287961 9.10
186.096476 61.562511 9.14
1.56135331 0.768001 4.13
-28.947564 10.104355 8.21
25.2121153 8.6077178 8.58
-3.5025523 1.7280995 4.11
-1.2619774 0.4359331 8.38
-1.8034997 0.6076278 8.81
Nparm DF L-R ChiSquare Prob>ChiSq
1 1 6.09880179 0.0135*
1 1 151.868665 <.0001*
2 2 111.379472 <.0001*
1 1 29.1823898 <.0001*
1 1 49.204595 <.0001*
Prediction Profiler

| AvgAgeUnresolvedCustQuestionsAtCoding | AvgWeeklyInPersonMeetings | ReqtsElicitationMethod | UnresolvedCustQuestionsAtCoding | UnresolvedDevQuestionsAtCoding |
|---|---|---|---|---|
| 17.39 | 0.11 | 1.00 | 22.63 | 36.33 |
| 17.39 | 0.11 | 1.00 | 22.63 | 40.585 |
| 17.39 | 0.11 | 1.00 | 22.63 | 44.84 |
| 17.39 | 0.11 | 1.00 | 22.63 | 49.095 |
| 17.39 | 0.11 | 1.00 | 22.63 | 53.35 |
| 17.39 | 0.11 | 1.00 | 25.7125 | 36.33 |
| 17.39 | 0.11 | 1.00 | 25.7125 | 40.585 |
| 17.39 | 0.11 | 1.00 | 25.7125 | 44.84 |
| 17.39 | 0.11 | 1.00 | 25.7125 | 49.095 |
| 17.39 | 0.11 | 1.00 | 25.7125 | 53.35 |
| 17.39 | 0.11 | 1.00 | 28.795 | 36.33 |
| 17.39 | 0.11 | 1.00 | 28.795 | 40.585 |
| 17.39 | 0.11 | 1.00 | 28.795 | 44.84 |
| 17.39 | 0.11 | 1.00 | 28.795 | 49.095 |
| 17.39 | 0.11 | 1.00 | 28.795 | 53.35 |
| 17.39 | 0.11 | 1.00 | 31.8775 | 36.33 |
| 17.39 | 0.11 | 1.00 | 31.8775 | 40.585 |
| 17.39 | 0.11 | 1.00 | 31.8775 | 44.84 |
| 17.39 | 0.11 | 1.00 | 31.8775 | 49.095 |
| 17.39 | 0.11 | 1.00 | 31.8775 | 53.35 |

| | Probability(Predicted CustomerSat=1) | Probability(Predicted CustomerSat=2) | Probability(Predicted CustomerSat=3) | Probability(Predicted CustomerSat=4) | Probability(Predicted CustomerSat=5) |
|---|---|---|---|---|---|
| 1 | 0.99995809 | 4.19142e-5 | 0 | 0 | 0 |
| 2 | 0.91728006 | 0.08271994 | 0 | 0 | 0 |
| 3 | 0.00512779 | 0.99487221 | 0 | 0 | 0 |
| 4 | 2.39571e-6 | 0.9999976 | 5.3291e-15 | 0 | 0 |
| 5 | 1.11354e-9 | 1 | 1.1385e-11 | 0 | 0 |
| 6 | 0.99795395 | 0.00204605 | 0 | 0 | 0 |
| 7 | 0.18480994 | 0.81519006 | 0 | 0 | 0 |
| 8 | 0.00010536 | 0.99989464 | 2.2204e-16 | 0 | 0 |
| 9 | 4.89789e-8 | 0.99999995 | 2.589e-13 | 0 | 0 |
| 10 | 2.2766e-11 | 1 | 5.5686e-10 | 0 | 0 |
| 11 | 0.90885654 | 0.09114346 | 0 | 0 | 0 |
| 12 | 0.00461353 | 0.99538647 | 0 | 0 | 0 |
| 13 | 2.15433e-6 | 0.99999785 | 5.9952e-15 | 0 | 0 |
| 14 | 1.00135e-9 | 1 | 1.266e-11 | 0 | 0 |
| 15 | 4.6543e-13 | 0.99999997 | 2.72379e-8 | 0 | 0 |
| 16 | 0.16934269 | 0.83065731 | 0 | 0 | 0 |
| 17 | 0.00009475 | 0.99990525 | 2.2204e-16 | 0 | 0 |
| 18 | 4.4044e-8 | 0.99999996 | 2.8777e-13 | 0 | 0 |
| 19 | 2.0472e-11 | 1 | 6.1926e-10 | 0 | 0 |
| 20 | 9.5155e-15 | 0.99999867 | 1.33229e-6 | 0 | 0 |
| 21 | 0.00415062 | 0.99584938 | 0 | 0 | 0 |



# Questions?

# Contact Information

**Robert W. Stoddard**

Email: rws@sei.cmu.edu

**Dave Zubrow**

Email: dz@sei.cmu.edu

**U.S. mail:**

Software Engineering Institute

Customer Relations

4500 Fifth Avenue

Pittsburgh, PA 15213-2612

USA

**World Wide Web:**

www.sei.cmu.edu

www.sei.cmu.edu/contact.html

**Customer Relations**

Email: customer-relations@sei.cmu.edu

Telephone: +1 412-268-5800

**SEI Phone:** +1 412-268-5800

**SEI Fax:** +1 412-268-6257